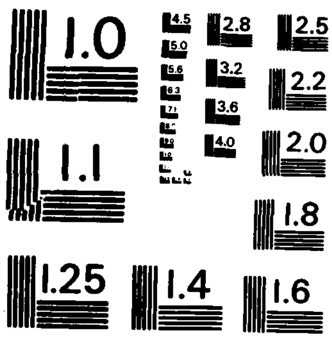
AD-A163 515 ORTHOGONAL REGRESSION M-ESTIMATORS(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS R H ZAMAR DEC 85 TR-62 N00014-84-C-0169 1/1

UNCLASSIFIED F/G 12/1 NL


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

ORTHOGONAL REGRESSION M-ESTIMATORS

by

Ruben H. Zamar

TECHNICAL REPORT No. 62

December 1985

Department of Statistics
University of Washington
Seattle, Washington 98195

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 62	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Orthogonal regression M-estimators		5. TYPE OF REPORT & PERIOD COVERED TR 12/1/84 - 11/30/85
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ruben H. Zamar		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, WA 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 661-003
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N 63374 1107 N.E. 45th Street Seattle, WA 98105		12. REPORT DATE December 1985
		13. NUMBER OF PAGES 52
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

M-estimators, orthogonal regression, qualitative robustness

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

M-estimators in the orthogonal regression set-up are defined in order to produce a robust alternative to the method of classical orthogonal regression. Asymptotic qualitative robustness of some orthogonal M-estimators is proved. The influence curve of these estimators is computed and some additional asymptotic theory is presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Orthogonal Regression M-Estimators

*Ruben H. Zamar**
University of Washington

ABSTRACT

M-estimators in the orthogonal regression set-up are defined in order to produce a robust alternative to the method of classical orthogonal regression. Asymptotic qualitative robustness of some orthogonal M-estimators is proved. The influence curve of these estimators is computed and some additional asymptotic theory is presented.

*This research was partially funded by the Office of Naval Research under contract no. N00014-84-C-0169, by the National Science Foundation, grant no. MCS 83-01807, and by CNPq (Brazil).

1. Introduction

The error-in-variables model and the method of orthogonal regression have received a great deal of attention in the statistics literature (see, for example, Anderson (1984) and Moran (1971) for extensive reviews and references). The method of orthogonal regression enjoys the following properties:

(P1) It is the ML procedure at the Gaussian error-in-variables model.

(P2) It gives a symmetric treatment to all the variables under study.

(P3) It allows for additional information about the variances of the observational errors (which might be available) to be used in order to obtain consistent estimators of the regression coefficients.

These are features which may render orthogonal regression more or less attractive, depending on the situation at hand.

A major disadvantage of orthogonal regression is that, like ordinary least squares regression, it lacks robustness: it is not resistant to outliers (see Brown, 1982), it lacks efficiency robustness its influence curve is unbounded (see Kelly, 1984), and its breakdown point is $1/n$ (see Hampel, 1971; Huber, 1981).

This paper studies a class of robust alternatives to orthogonal regression estimates, namely orthogonal regression M-estimates (ORM's). ORM-estimators have the property that they are MLE's for appropriate *non-Gaussian* error-in-variables models. They also enjoy properties (P2) and (P3).

In section 2, ORM-estimators are defined in the context of both symmetric and asymmetric models, the latter being the usual regression model with response and carriers. The existence and uniqueness of ORM-estimators is

established, under suitable assumptions, in section 3. Section 4 is devoted to ORM-estimator influence curves and some asymptotic bias calculations. Finally, section 5 treats the following asymptotic properties of ORM-estimators for both the structural and functional models: consistency, asymptotic normality and qualitative robustness.

2. Orthogonal Regression M-Estimators

Given a unit vector $\mathbf{a}$ in $\mathbb{R}^{p+1}$ and a real number b, the hyperplane $H(\mathbf{a},b)$ is defined as

$$H \;=\; H(\mathbf{a},b) \;=\; \{\mathbf{x}\in \mathbb{R}^{p+1}: \mathbf{a}'\mathbf{x} = b\}\ . \tag{2.1}$$

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of n data points in $\mathbb{R}^{p+1}$. The method of orthogonal regression (OR) consists of solving the following minimization problem:

$$\min_{\|\mathbf{a}\|=1, b\in\mathbb{R}} \sum_{i=1}^{n} d_i^2(\mathbf{x}_i, H) \;=\; \sum_{i=1}^{n} (\mathbf{a}'\mathbf{x}_i - b)^2 \tag{2.2}$$

where

$$d(\mathbf{x},H) \;=\; |\mathbf{a}'\mathbf{x} - b| \;=\; \min_{\mathbf{y}\in H} \|\mathbf{x} - \mathbf{y}\| \tag{2.3}$$

is the perpendicular (Euclidean) distance between $\mathbf{x} = (x_0, x_1, \ldots, x_p)'$ and H.

This view of orthogonal regression highlights the symmetry of the method, that is, no particular coordinate x_i of $\mathbf{x}$ is regarded as a response, with the remaining coordinates of $\mathbf{x}$ being predictors. However, orthogonal regression is often treated from the viewpoint of a linear model with coefficients $\beta_0, \beta_1, \ldots, \beta_p$ corresponding to an intercept and p predictor (or carrier) variables $x_1, \ldots, x_p$. Taking x_0 as the response, our symmetric form of orthogonal regression (2.2)-(2.3) can be cast in terms of the linear model coefficients as follows. Let

$$\begin{aligned} \boldsymbol{\beta} &= (\beta_0, \beta_1, \ldots, \beta_p)\,, \quad \delta = \left[1 + \sum_{i=1}^{p} \beta_i^2\right]^{-1/2} \\ \mathbf{a} &= \delta(1, -\beta_1, \ldots, -\beta_p)\,, \quad b = \delta\beta_0\ . \end{aligned} \tag{2.4}$$

Then

$$d^2(\mathbf{x},H) \;=\; \delta^2(x_0 - \beta_0 - \beta_1 x_1 - \cdots - \beta_p x_p)^2 \tag{2.5}$$

and solving the minimization problem

$$\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} [\delta(x_{0i} - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_p x_{pi})]^2 \tag{2.6}$$

is *almost* equivalent to solving (2.2). In fact the two solutions will be equivalent when (2.2) yields $\hat{a}_0 \neq 0$. For if $\hat{a}_0 \neq 0$, we can express a solution $\hat{\mathbf{a}} = (\hat{a}_0, \ldots, \hat{a}_p)$, $\|\hat{\mathbf{a}}\| = 1$ in the form (2.4) by letting $\hat{\beta}_i = -\hat{a}_i / \hat{a}_0$, $i = 1 \cdots p$ and $\hat{\delta} = (1 + \sum \hat{\beta}_i^2)^{-1/2}$. Note that a solution to (2.2) with $\hat{a}_0 = 0$ corresponds to $\beta_1 = \infty$ in the case $p = 1$, and to a hyperplane which is parallel to the x_0 direction in the general case. Of course data sets producing an estimate $\hat{\mathbf{a}}$ with $\hat{a}_0 = 0$ will have probability zero under any "reasonable" distribution model for $\mathbf{x}$.

As is well known, problem (2.6) yields the Gaussian MLE for the following error-in-variables (EV) model

$$X_0 = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{2.7}$$

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\varepsilon} \tag{2.7'}$$

where $\mathbf{x}' = (x_0, x_1, \ldots, x_p)$, $\mathbf{X}' = (X_0, X_1, \ldots, X_p)$, and $\boldsymbol{\varepsilon}' = (\varepsilon_0, \ldots, \varepsilon_p)$ is $N(\mathbf{0}, \sigma^2 \mathbf{I})$. When we take n observations on the above model we write

$$\mathbf{x}_i = \mathbf{X}_i + \boldsymbol{\varepsilon}_i \tag{2.8}$$

where $\mathbf{x}'_i = (x_{0i}, x_{1i}, \ldots, x_{pi})$, $\mathbf{X}_i = (X_{0i}, X_{1i}, \ldots, X_{pi})$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{0i}, \varepsilon_{1i}, \ldots, \varepsilon_{pi})$ for $i = 1, \ldots, n$. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are random iid vectors we have a *structural* error-in-variables model. On the other hand, if $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are fixed (non-random) but unknown *incidental* parameters we have a *functional* error-in-variables model.

If $\mathbf{a}_0$ is a unit vector in $\mathbb{R}^{p+1}$ and $b_0 \in \mathbb{R}^1$, the symmetric formulation of the model (2.7)-(2.7') is

$$\mathbf{a}_0 \mathbf{X} = b_0 \tag{2.9}$$

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\varepsilon} \tag{2.9'}$$

and (2.4) expresses the relationship between $(\mathbf{a}_0, b_0)$ and $\boldsymbol{\beta} = (\beta_1, \beta_1, \ldots, \beta_p)$. The noise-free part of the error-in-variables (EV) model can be represented geometrically as a linear variety $\mathbf{M} = \mathbf{m}_0 + \mathbf{V}$, where $\mathbf{V}$ is a p-dimensional subspace of $\mathbb{R}^{p+1}$ and $\mathbf{m}_0$ is a solution of either (2.7) or (2.9). This is illustrated in Figure 2.1 for $p = 1$, i.e., $X_0 = \beta_0 + \beta_1 X_1$.

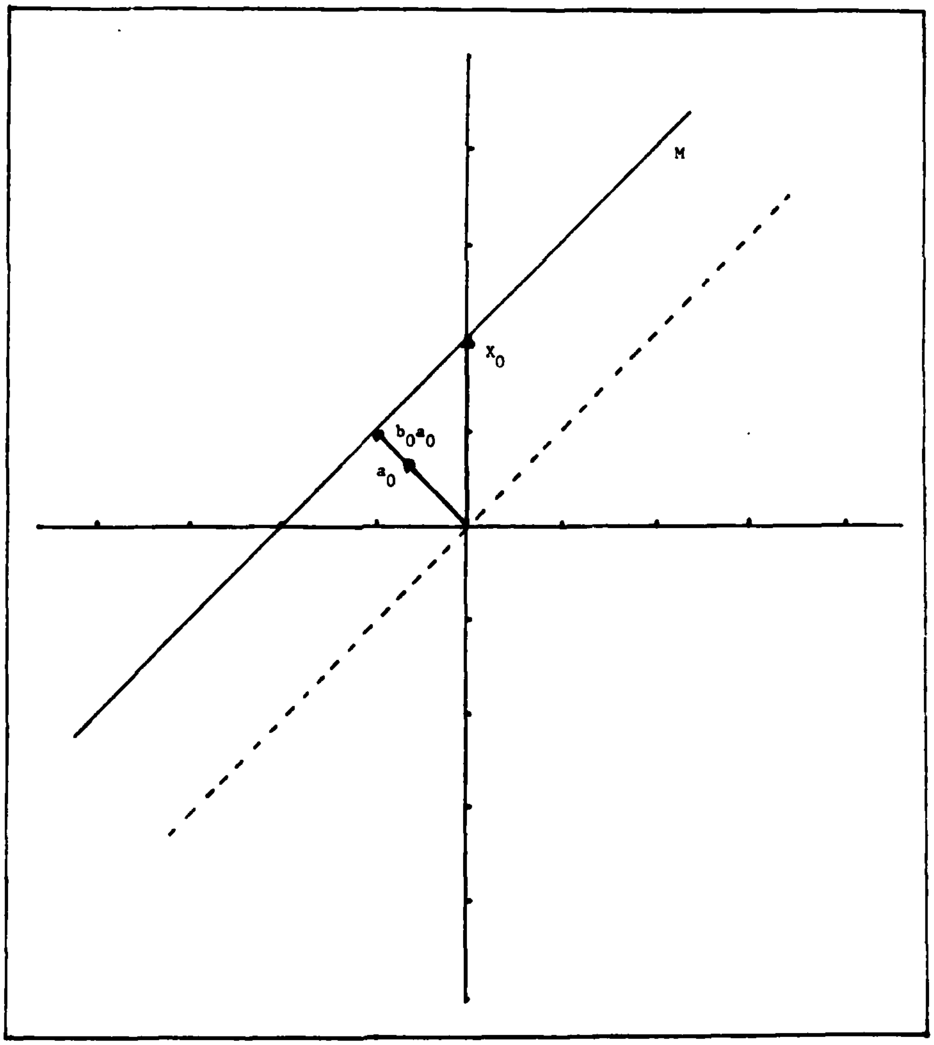


FIGURE 2.1

In the linear regression setup (2.7) $\mathbf{m}_0 = \mathbf{m}_0^r = (\beta_0, 0, \ldots, 0)$, and in the symmetric setup (2.9) $\mathbf{m}_0 = \mathbf{m}_0^s = b_0\mathbf{a}_0$.

Further perusal of the linear regression formulation of the EV model leads to a nice geometric motivation for the definition of the orthogonal regression M-estimate **(ORM)**. First one verifies that

$$\begin{aligned}\mathbf{b}_1 &= \delta(\beta_1, 1, 0, \ldots, 0)\\ \mathbf{b}_2 &= \delta(\beta_2, 0, 1, \ldots, 0)\\ &\vdots\\ \mathbf{b}_p &= \delta(\beta_p, 0, 0, \ldots, 1)\ .\end{aligned} \tag{2.10}$$

is a (non-orthogonal) basis for $\mathbf{V}$. Since $\mathbf{a}_0$ is orthogonal to $\mathbf{V}$ any vector $\mathbf{x} \in \mathbb{R}^{p+1}$ can be written as

$$\mathbf{x} = \mathbf{m}_0^r + \sum_{k=1}^{p} z_k \mathbf{b}_k + e\,\mathbf{a}_0 \quad \forall\ z_1, \ldots, z_p \in \mathbb{R}' \tag{2.11}$$

where $\sum_{k=1}^{p} z_k \mathbf{b}_k \in \mathbf{V}$, $\mathbf{x}_0 + \sum_{k=1}^{p} z_k \mathbf{b}_k$ is the component of $\mathbf{x}$ which lies in the noise-free model (2.7), and $e\,\mathbf{a}_0$ is orthogonal to the noise-free model (2.7). The simple case $p = 1$ and $\beta_0 = 0$ is illustrated in Figure 2.2 below.

The coefficients $z_1, \ldots, z_p$ are easily determined by taking the inner product of $\mathbf{c}_k = (0, 0, \ldots, 1, \cdots 0)'$, with both sides of (2.11), where the "1" is in the $k + 1$st position of $\mathbf{c}_k$. This gives

$$z_k = \delta^{-1}(x_k + \delta\beta_k e)\ . \tag{2.12}$$

Finally e is determined by premultiplying both sides of (2.11) by $\mathbf{a}'$, which gives

$$e = \delta(x_0 - \beta_0 - \sum_{k=1}^{p} \beta_k x_k)\ . \tag{2.13}$$

It is straightforward to check that

$$\frac{\partial}{\partial\beta_0} e = -\delta\ , \quad \frac{\partial}{\partial\beta_k} e = -\delta^2 z_k \qquad (k = 1, \ldots, p)\ . \tag{2.14}$$

Then, since the minimization problem (2.6) can be expressed in terms of the

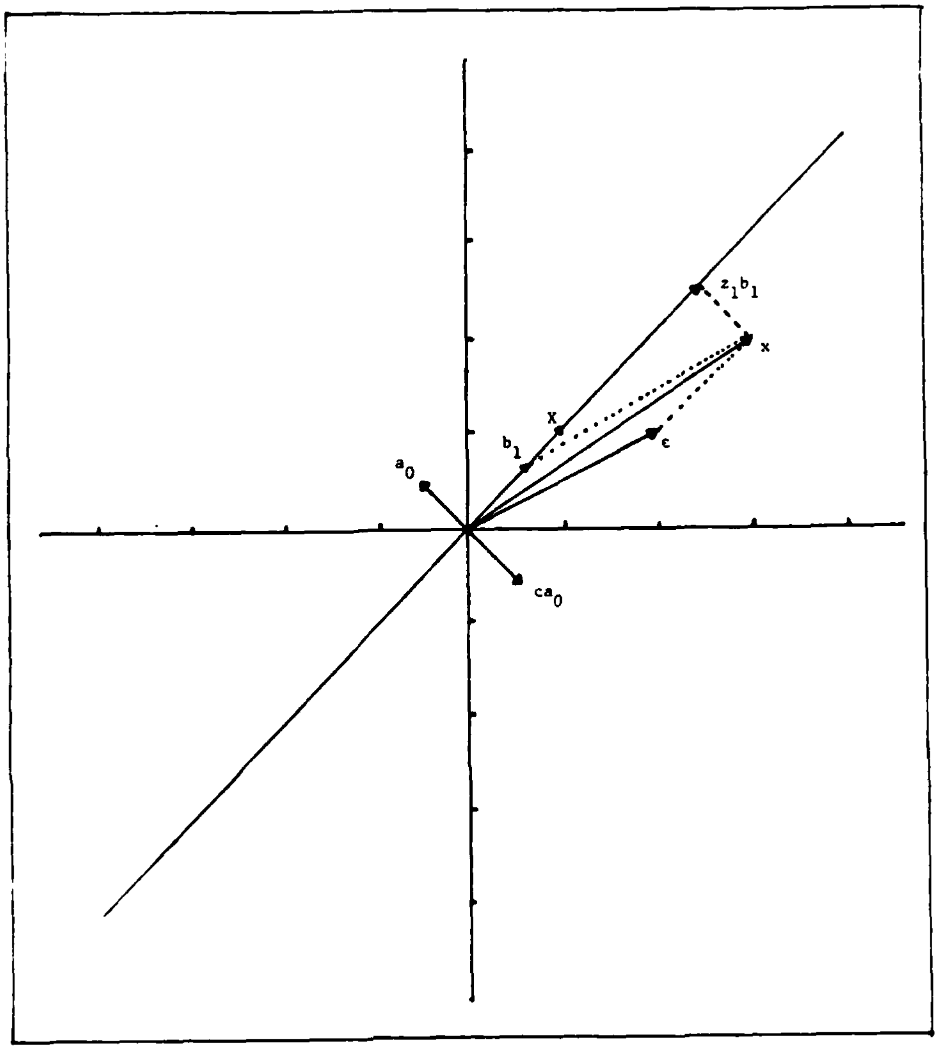


FIGURE 2.2

orthogonal errors $e_i = \delta(x_{0i} - \beta_0 - \sum_{k=1}^{p} \beta_k x_{ki})$, $i = 1, \ldots, n$, as

$$\frac{1}{n} \sum_{i=1}^{n} e_i^2 = \min_{\beta \in \mathbb{R}^{p+1}}! \tag{2.15}$$

differentiation gives the "orthogonal normal equations"

$$
\begin{aligned}
&\text{(a)} \quad \frac{1}{n}\sum_{i=1}^{n} e_i z_{ki} = 0 \\
&\text{(b)} \quad \frac{1}{n}\sum_{i=1}^{n} e_i = 0 \ .
\end{aligned}
\tag{2.16}
$$

This leads to a natural definition of an *orthogonal regression M-estimator:*

$$
\frac{1}{n}\sum_{i=1}^{n} \rho(e_i) = \min!
\tag{2.17}
$$

where ρ is a symmetric robustifying function. Differentiation with respect to $\boldsymbol{\beta}$ gives the following ORM estimating equations:

$$
\begin{aligned}
&\text{(a)} \quad \frac{1}{n}\sum_{i=1}^{n} \psi(e_i) z_{ki} = 0 \\
&\text{(b)} \quad \frac{1}{n}\sum_{i=1}^{n} \psi(e_i) = 0
\end{aligned}
\tag{2.18}
$$

where $\psi = \rho'$.

MLE MOTIVATION

There is also a maximum likelihood rationale for the above definition of ORM-estimators. Let

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 X_i + v_i \\
x_i &= X_i + u_i
\end{aligned}
\tag{2.19}
$$

be the usual error-in-variable model with all the classical assumptions in force except that the joint distribution of the error (u,v) is not Gaussian but of the form

$$
f(u,v) = K e^{-\rho(u^2+v^2)} \ .
\tag{2.20}
$$

The log-likelihood function for this model is

$$
l(\beta_0, \beta_1, X_1, \ldots, X_n) = n\log(K) - \sum^{n} \rho[(y_i - \beta_0 - \beta_1 X_i)^2 + (x_i - X_i)^2] \ .
\tag{2.21}
$$

Differentiation with respect to all the parameters (structural and incidental) gives

$$\frac{\partial l}{\partial \beta_1} = 0 \quad => \quad \frac{1}{n}\sum^{n} \psi[(y_i-\beta_0-\beta_1 X_i)^2 + (x_i-X_i)^2](y_i-\beta_0-\beta_1 X_i)X_i = 0 \tag{2.22a}$$

$$\frac{\partial l}{\partial \beta_0} = 0 \quad => \quad \frac{1}{n}\sum \psi[(y_i-\beta_0-\beta_1 X_i)^2 + (x_i-X_i)^2](y_i-\beta_0-\beta_1 X_i) = 0 \tag{2.22b}$$

$$\frac{\partial l}{\partial X_i} = 0 \quad => \quad \psi[(y_i-\beta_0-\beta_1 X_i)^2 + (x_i-X_i)^2][(y_i-\beta_0-\beta_1 X_i)\beta_1 + (x_i-X_i)] = 0 \ ,$$

$$(i = 1, \ldots, n) \ . \tag{2.22c}$$

Equations (2.22c) are all satisfied if we set

$$\hat{X}_i = \frac{x_i + \beta_1(y_i-\beta_0)}{1+\beta_1^2} \ , \qquad (i = 1, \ldots, n) \ . \tag{2.23}$$

Inserting (2.23) into (2.22a) and (2.22b) we get the following MLE estimating equations:

$$\begin{aligned} \frac{1}{n}\sum^{n} \tilde{\psi}\left[\frac{y_i-\beta_0-\beta_1 x_i}{(1+\beta_1^2)^{1/2}}\right]\left[\frac{x_i+\beta_1(y_i-\beta_0)}{(1+\beta_1^2)^{1/2}}\right] &= 0 \\ \frac{1}{n}\sum^{n} \tilde{\psi}\left[\frac{y_i-\beta_0-\beta_1 x_i}{(1+\beta_1^2)^{1/2}}\right] &= 0 \end{aligned} \tag{2.24}$$

where $\tilde{\psi}(t) = \psi(t^2)t$. But equations (2.24) are also the estimating equation of the following (ORM-estimator) optimization problem:

$$\min_{\beta_0, \beta_1} \frac{1}{n}\sum^{n} \tilde{\rho}\left[\frac{y_i-\beta_0-\beta_1 x_i}{(1+\beta_1^2)^{1/2}}\right] \tag{2.25}$$

where $\tilde{\rho}(t) = \frac{1}{2}\rho(t^2)$ and $\tilde{\psi}(t) = \tilde{\rho}'(t) = \psi(t^2)t$. This carries over to the general p case, as shown in the Appendix.

3. Existence and Uniqueness of ORM-Estimators

We now show that given a sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of observations in $\mathbf{R}^{p+1}$ the minimization problem

$$\min_{\|\mathbf{a}\|=1,\, b\in\mathbf{R}} \frac{1}{n} \sum \rho(\mathbf{a}'\mathbf{x}_i - b) \tag{3.1}$$

has at least one solution. In general the solution to (3.1) is not unique but we will show that appropriate side constraints can be imposed in order to force uniqueness. On the other hand we'll see later on that it is not necessary to force uniqueness in the context of asymptotics since all solutions to (3.1) will enjoy the same asymptotic behavior. We consider two cases.

MONOTONE ψ

For each unit vector $\mathbf{a}$ in $\mathbf{R}^p$ let

$$\mathrm{B}(\mathbf{a}) = \left\{ \hat{b} : \min_{b\in\mathbf{R}} \frac{1}{n} \sum \rho(\mathbf{a}\mathbf{x}_i - b) = \frac{1}{n} \sum \rho(\mathbf{a}'\mathbf{x}_i - \hat{b}) \right\} . \tag{3.2}$$

Monotonicity of $\psi = \rho'$ implies that $\rho(t) \to +\infty$ as $t \to \pm\infty$, and therefore it follows as in Huber (1964) that $\mathrm{B}(\mathbf{a})$ is a non-empty, bounded and closed interval.

Lemma 3.1: Let $\mathbf{a}, \mathbf{c}$ denote unit vectors in $\mathbf{R}^{p+1}$. For each $\varepsilon > 0$ there exists $\delta > 0$ such that if $\|\mathbf{a} - \mathbf{c}\| < \delta$ and $b_1 \in \mathrm{B}(\mathbf{a})$ then there exists $b_2 \in \mathrm{B}(\mathbf{c})$ with $|b_1 - b_2| < \varepsilon$.

Proof: See the Appendix.

Now we select a particular element of $B(\mathbf{a})$, namely

$$F(\mathbf{a}) = \min B(\mathbf{a}) . \tag{3.3}$$

Lemma 3.2: $F(\mathbf{a})$ is a uniformly continuous function of $\mathbf{a}$.

Proof: See the Appendix.

Lemma 3.3: Given $\mathbf{x}_1, \ldots, \mathbf{x}_n$, for all $\|\mathbf{a}\| = 1$ and $b \in \mathbf{R}$

$$\sum \rho(\mathbf{a}'\mathbf{x}_i - b) \geq \sum \rho(\mathbf{a}'\mathbf{x}_i - F(\mathbf{a})) . \tag{3.4}$$

Proof: If $b_0 = F(\mathbf{a})$ and $b \in \mathbf{R}$ then

$$\sum^n \rho(\mathbf{a}'\mathbf{x}_i - b) = \sum^n \rho[\mathbf{a}'\mathbf{x}_i - b_0 + (b_0 - b)] =$$

$$= \sum^n \rho(\mathbf{a}'\mathbf{x}_i - b_0) + (b_0 - b) \sum^n \psi[\mathbf{a}'\mathbf{x}_i - b_0 + \alpha_i(b_0 - b)] \tag{3.5}$$

where $0 \leq \alpha_i \leq 1$ for $i = 1, \ldots, n$. Since $t\psi(s+t) \geq t\psi(s)$ for all $s,t \in \mathbf{R}$ we have

$$\sum^n \rho(\mathbf{a}'\mathbf{x}_i - b) \geq \sum^n \rho(\mathbf{a}'\mathbf{x}_i - b_0) + (b_0 - b) \sum \psi(\mathbf{a}'\mathbf{x} - b_0)$$

$$\geq \sum^n \rho(\mathbf{a}'\mathbf{x}_i - b_0) . \tag{3.6}$$

The above Lemmas yield the following:

Proposition 3.1: The minimization problem (3.1) has at least one solution.

Proof: Since the right hand side of (3.4) is a continuous function of $\mathbf{a}$ it achieves its minimum at some vector $\hat{\mathbf{a}}$, say. If $\hat{b} = F(\hat{\mathbf{a}})$ then $(\hat{\mathbf{a}}, \hat{b})$ is a solution to (3.4).

BOUNDED ρ

We note that when ρ is bounded the ψ-function in the ORM estimating equation is of the redescending type. It will be shown in the next section that under certain conditions an ORM-estimator with a bounded loss function ρ is robust in a precise technical sense.

Proposition 3.2: Let ρ be a non-negative, even function which is non-decreasing on $[0,\infty)$ and such that

$$0 < \lim_{|t|\to\infty} \rho(t) = M < \infty . \tag{3.7}$$

Then (3.1) has at least one solution.

Proof: Let $0 < \delta < M/n$ and $t_0 > 0$ such that

$$|t| > t_0 \quad => \quad \rho(t) > M - \delta . \tag{3.8}$$

Let $K = \max_{1\le i\le n} \|\mathbf{x}_i\|$. If $|b| > t_0+K$ then

$$|\mathbf{a}'\mathbf{x}_i - b| \ \geq\ |b| - |\mathbf{a}'\mathbf{x}_i| \ \geq\ |b| - K \ >\ t_0 . \tag{3.9}$$

Consequently if $|b| > t_0+K$, for all $\|\mathbf{a}\| = 1$ and all $i = 1, \ldots, n$

$$\rho(\mathbf{a}'\mathbf{x}_i - b) > M - \delta \tag{3.10}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{a}'\mathbf{x}_i - b) > M - \delta . \tag{3.11}$$

Since $\frac{1}{n}\sum^{n} \rho(\mathbf{a}'\mathbf{x}_i - b)$ is a continuous function of $(\mathbf{a},b)$ it achieves its minimum on the compact set

$$\mathbf{A} = \{(\mathbf{a},b) : \|\mathbf{a}\| = 1 ,\ \ |b| \le t_0 + K\} . \tag{3.12}$$

Let $\mathbf{c}$ be such that $\mathbf{c}'\mathbf{x}_1 = 0$ and $\|\mathbf{c}\| = 1$. Now $(\mathbf{c},0) \in \mathbf{A}$ so that by definition of δ

$$\min_{(\mathbf{a},b)\in A} \frac{1}{n}\sum \rho(\mathbf{a}'\mathbf{x}_i - b) \le \frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{c}'\mathbf{x}_i) \le \frac{n-1}{n} M < M - \delta$$

$$\le \inf_{\|\mathbf{a}\|=1,\, |b|>K+t_0} \frac{1}{n}\sum \rho(\mathbf{a}'\mathbf{x}_i - b) \;. \qquad (3.12')$$

Therefore $\frac{1}{n}\sum^{n}\rho(\mathbf{a}'\mathbf{x}_i - b)$ achieves its minimum when $(\mathbf{a},b)$ ranges on $\{(\mathbf{a},b) : \|\mathbf{a}\| = 1,\ b \in \mathbf{R}\}$ and the minimum can be achieved only at elements of A.

FORCING UNIQUENESS

So far there is no assurance of uniqueness of solutions in either of the two cases considered. However, when the solution set is not a singleton, it is possible to select a particular solution $\hat{\theta}_n = (\hat{\mathbf{a}}_n, \hat{b}_n)$ from the (compact) solution set in a systematic way, as we now show. Let

$$\mathrm{B} = \left\{ (\mathbf{a},b) : \frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{a}'\mathbf{x}_i - b) = \min_{\|\mathbf{c}\|=1,\, d\in\mathbf{R}} \frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{c}'\mathbf{x}_i - d) \right\} . \qquad (3.13)$$

Among all the minimizing pairs $(\mathbf{a},b)$ we take only those with the smallest possible non-negative b-component to form the set B_0, that is

$$\begin{aligned} \bar{b} &= \inf\{b : (\mathbf{a},b) \in \mathrm{B}\} \\ \mathrm{B}_0 &= \{(\mathbf{a},b) \in \mathrm{B} : b = \bar{b}\} \end{aligned} \qquad (3.14)$$

Now let $\{\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}\}$ be the canonical basis of $\mathbf{R}^{p+1}$ and for $k = 1, \ldots, p+1$ let

$$\mathrm{B}_k = \left\{ (\mathbf{a},b) \in \mathrm{B}_{k-1} : \|\mathbf{a} - \mathbf{u}_k\| \le \|\mathbf{c} - \mathbf{u}_k\| \;, \quad \forall\ (\mathbf{c},b) \in \mathrm{B}_{k-1} \right\} . \qquad (3.15)$$

That is, among all the pairs $(\mathbf{c},b)$ in B_{k-1} we take only those which are nearest to $\mathbf{u}_k$ to form the set B_k. Finally the following lemma provides us with a uniquely defined element of B which we shall label as $\hat{\theta}_n$.

Lemma 3.4: $\mathbf{B}_{p+1}$ defined above contains exactly one element.

Proof: Assume that $(\mathbf{a},\bar{b})$ and $(\mathbf{c},\bar{b})$ are in $\mathbf{B}_{p+1}$. Then $\|\mathbf{a}\| = \|\mathbf{c}\| = 1$ and

$$\|\mathbf{a}-\mathbf{u}_k\|^2 = \|\mathbf{c}-\mathbf{u}_k\|^2 \ , \qquad k = 1, \ldots, p+1 \ . \tag{3.16}$$

But if $\mathbf{a} = (a_1, \ldots, a_{p+1})'$ and $\mathbf{c} = (c_1, \ldots, c_{p+1})$ then

$$\|\mathbf{a}-\mathbf{u}_k\|^2 = (a_k-1)^2 + \sum_{j\neq k} a_j^2 = 1 - 2a_k + \sum_{j=1}^{p+1} a_j^2 = 2(1-a_k) \tag{3.17}$$

and the lemma follows since $2(1-a_k) = \|\mathbf{a}-u_k\|^2 = \|\mathbf{c}-\mathbf{u}_k\|^2 = 2(1-c_k)$ implies that $\mathbf{a}=\mathbf{c}$.

4. Influence Curve and Asymptotic Bias of ORM-Estimators

4.1 INFLUENCE CURVE

Hampel's (1974) *influence curve* is one of the most useful heuristic tools in robustness (cf. Huber, 1981). The study of its shape enables one to visualize the way a small fraction of contaminated data can modify the asymptotic behavior of a certain estimator. In addition it yields the asymptotic variance of the estimator. In this section we compute the influence curve of the ORM-estimator and study its behavior in more detail for some particular models.

In order to define the influence curve of the ORM-estimator we assume a nominally Gaussian distribution for the EV model (2.7)-(2.7'). That is, we assume that

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\varepsilon} \tag{4.1}$$

where $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are independent, $\mathbf{X} \sim N_{p+1}(\boldsymbol{\mu},\Sigma)$, $\boldsymbol{\varepsilon} \sim N_{p+1}(\mathbf{0},\sigma^2\mathbf{I})$, and the random vector $\mathbf{X} = (X_0, X_1, \ldots, X_p)'$ a.s. satisfies the linear model

$$X_0 = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \ . \tag{4.2}$$

Thus $F = N(\boldsymbol{\mu},\Sigma+\sigma^2 I)$ is the distribution function of $\mathbf{x}$. Now let δ_{ζ} be a point mass located at $\boldsymbol{\zeta} \in \mathbb{R}^{p+1}$, and for $0 < \varepsilon < 1$, let

$$G_{\varepsilon,\zeta} = (1-\varepsilon)F + \varepsilon\delta_{\zeta} \ . \tag{4.3}$$

The influence curve of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)$ is defined as

$$\mathrm{IC}(\zeta) = \mathrm{IC}(\zeta; \hat{\boldsymbol{\beta}}, F) = \lim_{\varepsilon \to 0} \frac{\hat{\boldsymbol{\beta}}(G_{\varepsilon,\zeta}) - \hat{\boldsymbol{\beta}}(F)}{\varepsilon} \tag{4.4}$$

provided the limit exists. Here the estimator $\hat{\boldsymbol{\beta}}$ is regarded as a $\mathbb{R}^{p+1}$-valued functional defined on some subset of distribution functions.

We compute the influence curve of the ORM-estimator in the usual way. Namely, we differentiate the population (or asymptotic) estimating equations of the M-estimator under $G_{\varepsilon,\zeta}$ with respect to ε, and evaluate the derivative at $\varepsilon = 0$. This produces a system of linear equations which determine IC(ζ).

The (population) estimating equations of the ORM-estimator under $G_{\varepsilon,\zeta}$ are

$$A_k(\varepsilon) = (1-\varepsilon)E\{\psi(e)z_k\} + \varepsilon\psi[e(\zeta,\beta)]z_k(\zeta,\beta) = 0 \tag{4.5}$$

for $k = 1, \ldots, p$ and

$$A_0(\varepsilon) = (1-\varepsilon)E\{\psi(e)\} + \varepsilon\psi[e(\zeta,\beta)] = 0 \ . \tag{4.6}$$

Here $z_k = z_k(\zeta,\beta)$, $k = 1, \ldots, p$ and $e = e(\zeta,\beta)$ are defined on (2.12) and (2.13). Also $\beta = \beta[(1-\varepsilon)F + \varepsilon\delta_\zeta]$ is the asymptotic value of this parameter under $G_{\varepsilon,\zeta}$. In addition $E\{\psi(e)z_k\}$ and $E\{\psi(e)\}$ stand for the more explicit notation $E_F\{\psi[e(\mathbf{x},\beta)]z(\mathbf{x},\beta)\}$ and $E_F\{\psi[e(\mathbf{x},\beta)]\}$, respectively.

Differentiation of (4.5) and (4.6) at $\varepsilon = 0$ can be carried out with the help of (2.14) and the following easily-derived relations:

$$\frac{\partial}{\partial\beta_k} z_j = \begin{cases} -\delta\beta_j & k = 0 \\ e(1-\delta^2\beta_j^2) & k = j \\ \delta^2(\beta_k z_j - \beta_j z_k - \beta_k\beta_j e) & k \neq j\ ,\ k \neq 0\ . \end{cases} \tag{4.7}$$

for $j = 1, \ldots, p$. The resulting linear equations are

$$\mathbf{M}\, IC(\zeta;\beta) = \gamma(\zeta,\beta) \tag{4.8}$$

where

$$\mathrm{IC}(\zeta,\beta) = (IC_0(\zeta,\beta), \ldots, IC_p(\zeta,\beta))'$$

$$\gamma(\zeta,\beta) = \psi[e(\zeta,\beta)](1, z_1(\zeta,\beta), \ldots, z_p(\zeta,\beta))' \tag{4.9}$$

and $\mathbf{M}$ is a $(p+1)\times(p+1)$ matrix with elements

$$m_{0j} = \begin{cases} \delta E\{\psi'(e)\} & j = 0 \\ \delta^2 E\{\psi'(e)z_j\} & j = 1, \ldots, p \end{cases} \tag{4.10}$$

and

$$m_{kj} = \begin{cases} \delta E\{\psi'(e)z_k\} & j = 0 \\ \delta^2 E\{\psi'(e)z_k^2\} - E\{\psi(e)e\}(1-\delta^2\beta_k^2) & j = k \\ \delta^2 E\{\psi'(e)z_k z_j\} + \delta^2 E\{\psi(e)e\}\beta_k\beta_j & j \neq k,0 \end{cases} \tag{4.11}$$

for $k = 1, \ldots, p$.

The entries of **M** are in general rather complicated, which makes it difficult to give an explicit expression for $\mathbf{M}^{-1}$ (and hence IC (ζ,β)) in the general case. However, this is easy to do for the simple case where $p = 1$ and $E\{X_1\} = 0$. In this case e and z_1 are independent normal variables with zero means, $m_{01} = m_{10} = 0$, and one can check that $(1-\delta^2\beta_1^2) = \delta^2 = (1+\beta_1^2)^{-1}$. Therefore,

$$IC_0(\zeta,\beta) = \frac{(1+\beta_1^2)^{1/2}}{E\{\psi'(e)\}}\,\psi[e(\zeta,\beta)] \tag{4.12}$$

and

$$IC_1(\zeta,\beta) = \frac{1+\beta_1^2}{E\{\psi'(e)z_1^2 - \psi(e)e\}}\,\psi[e(\zeta,\beta)]z_1(\zeta,\beta)\ . \tag{4.13}$$

Here the intercept component $IC(\zeta,\beta_0)$ is independent of z_1 and is bounded, while the slope component $IC(\zeta,\beta_1)$ is unbounded.

REMARK: OR corresponds to the particular case $\psi(t) = t$ and $\psi'(t) = 1$ (cf. Kelly, 1984, eqs. (2.6) and (2.7)).

It is clear from (4.8) and (4.9) that the elements of IC (ζ,β) are in general unbounded. Nevertheless if ψ is fully redescending (e.g., if ψ is Tukey's bisquare function), then $IC(\zeta,\beta)$ is unbounded in a nice way: it can be arbitrarily large only when $e(\zeta,\beta)$, the departure of ζ from the true error-free model, is small. (This is also the behavior of an ordinary M-estimate using redescending ψ for the classical linear model setup.)

To illustrate this consider the simple case $p = 1$ and $\beta_0 = \beta_1 = 0$, and with ψ fully redescending. For each $\zeta_1 \neq 0$ and $\alpha > 0$ let

$$\mathbf{A}_\alpha(\zeta) \;=\; \{(\zeta_1, \zeta_2):\ \zeta_1 = \zeta,\ |\zeta_2/\zeta| \geq \alpha\} \ .$$

For $\alpha = 0$ let

$$\mathbf{A}_\alpha(0) \;=\; \{(0, \zeta_2):\ -\infty < \zeta < \infty\}$$

$$\mathbf{A}_\alpha \;=\; \bigcup_{\zeta \in \mathbf{R}} \mathbf{A}_\alpha(\zeta)$$

and

$$g_\alpha(\zeta_1) \;=\; \sup_{\zeta_2 \in \mathbf{A}_\alpha(\zeta_1)} IC_1(\zeta, \beta) \ .$$

Notice that points in $\mathbf{A}_\alpha^c = \mathbb{R}^2 - \mathbf{A}_\alpha$ are "nice" (for small α) in the sense that they belong to lines through the origin with slope less than α (see Figure 4.1 below).

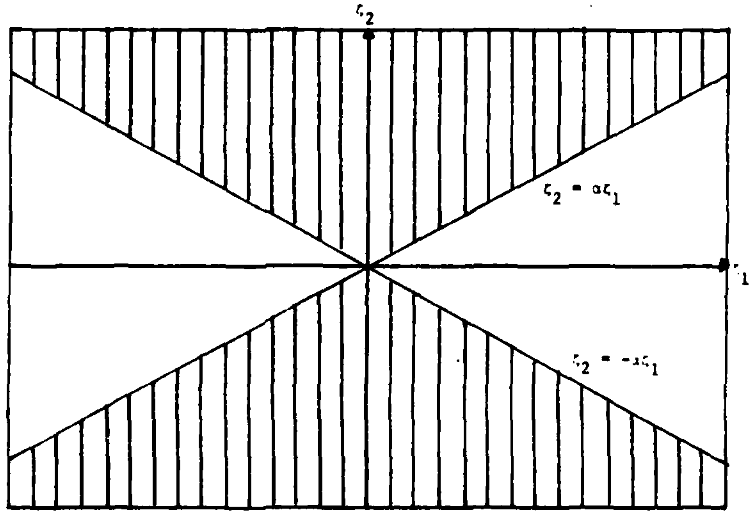


FIGURE 4.1

In Figure 4.2 below are presented plots of $g_\alpha(\zeta_1)$ versus ζ_1 for various values of α. It shows that the maximum influence of points outside $\mathbf{A}_\alpha$ is finite for all $\alpha > 0$ and that this uniformly decreases as α gets larger.

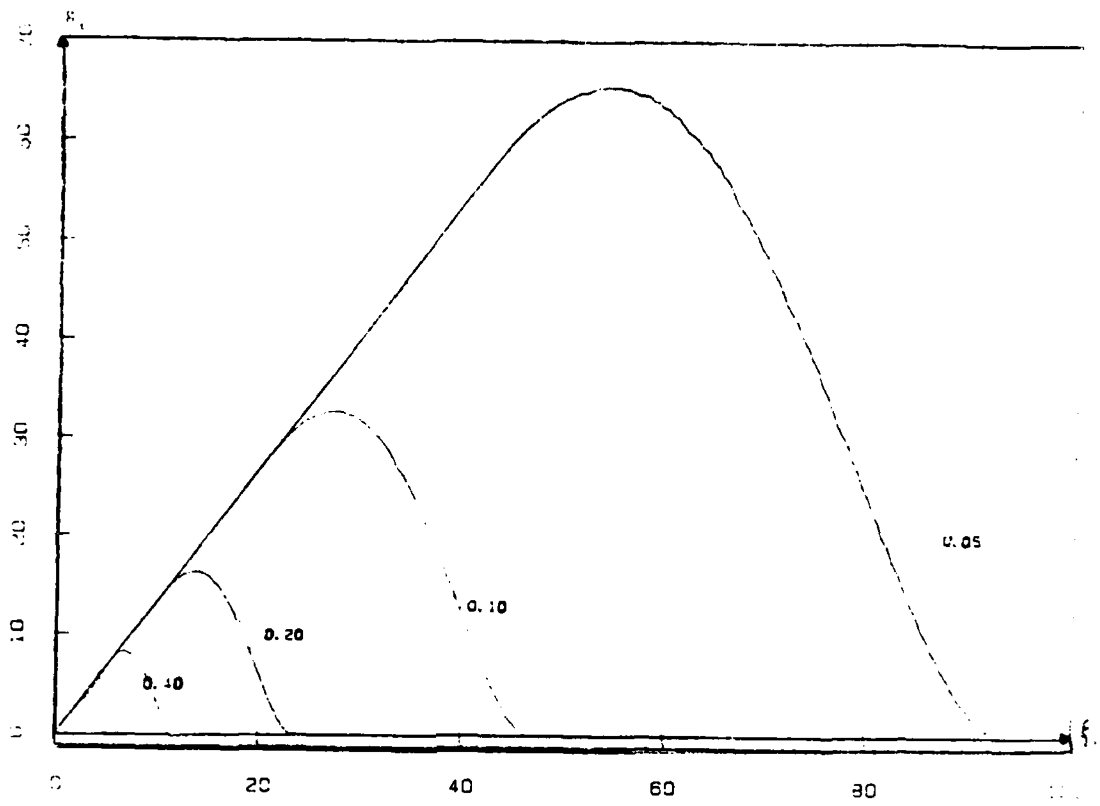


FIGURE 4.2

Another case, with general p, which yields an IC whose intercept component is independent of $z_1, \ldots, z_p$ arises when $\beta_1 = \cdots = \beta_p = E\{X_1\} = \cdots = E\{X_p\}$ = 0. For in this case it is straightforward to see that $m_{0j} = 0$ ($j = 1, \ldots, p$), $m_{k0} = 0$ ($k = 1, \ldots, p$), $m_{00} = E\{\psi'(e)\}$ and for $k = 1, \ldots, m$

$$m_{kj} = \begin{cases} E\{\psi'(e)\}E\{z_k^2\} - E\{\psi(e)e\} & j = k \\ E\{\psi'(e)\}E\{z_j z_k\} & j \neq k \end{cases} \tag{4.14}$$

Now let

$$\sigma = \begin{bmatrix} \sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} m_{00} & \mathbf{M}_{01} \\ \mathbf{M}_{10} & \mathbf{M}_{11} \end{bmatrix} \tag{4.15}$$

where σ_{00} and m_{00} are of order 1×1, and Σ_{11} and M_{11} are of order $p \times p$. Then we have $m_{00} = E\{\psi'(e)\}$, $\mathbf{M}_{10} = \mathbf{M}'_{10} = \mathbf{0}$, and

$$\mathbf{M}_{11} = E\{\psi'(e)\}\Sigma_{11} + (\sigma^2 - E\{\psi(e)e\})\mathbf{I}_p \ . \tag{4.16}$$

If $E\{\psi'(e)\} \geq 0$ and $\sigma^2 - E\{\psi(e)e\} > 0$ (as is generally the case), then $\mathbf{M}_{11}$ is positive definite and invertible.

5. Asymptotic Theory

In this section the structural and functional error-in-variables models introduced in section 2 ((2.9) and (2.9')) are considered in order to establish some asymptotic properties of the **ORM**-estimator.

It is well known (Kendall and Stuart, 1979) that when the covariance matrix of ε is known (up to a constant factor), the classical orthogonal regression estimator is consistent under both the structural and the functional EV models.

Subsection 5.1 is devoted to the study of the asymptotic properties of ORM-estimators under the structural model. Consistency (Theorem 5.1), asymptotic normality (Theorem 5.2), and qualitative robustness (Theorem 5.3) are established. The asymptotic bias of a particular ORM-estimator under some contaminated Gaussian EV models is numerically computed and presented in Table 5.1. Subsection 5.2 is concerned with the asymptotic properties of ORM-estimators under the functional EV model. Consistency (Theorem 5.4) and asymptotic normality (subsection 5.2.2) are treated here.

5.1 STRUCTURAL CASE

5.1.1 Consistency

First we show (Theorem 5.1) that under fairly weak assumptions, an ORM-estimate converges to its asymptotic (population) value. Then, Corollary 5.1 establishes the consistency of an ORM-estimator (convergence to the true parameter in the EV-model) under additional distributional assumptions, including that $\mathbf{A}\boldsymbol{\varepsilon}$ is spherically symmetric for some known, non-singular matrix $\mathbf{A}$.

Theorem 5.1: Assume that the loss function ρ is continuous, non-negative, and non-decreasing on $[0,\infty)$. Suppose that there exists a unit vector $\mathbf{a}_1$ and a real number b_1 such that

(i) $E_F\{\rho(\mathbf{a}'_1\mathbf{x}-b_1)\} < \lim_{t\to\infty}\rho(t)$.

(ii) $(\mathbf{a}_1,b_1)$ (strictly up to the sign) minimizes $E_F\{\rho(\mathbf{a}'\mathbf{x}-b)\}$ among all unit vectors $\mathbf{a}$ and real numbers b.

If $\hat{\boldsymbol{\theta}}_n = (\hat{\mathbf{a}}_n,\hat{b}_n)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\hat{\mathbf{a}}_n{}'\mathbf{x}_i-\hat{b}_n) - \inf_{\|\mathbf{a}\|=1,b\in\mathbf{R}}\frac{1}{n}\sum_{i=1}^{n}\rho(\mathbf{a}'\mathbf{x}_i-b) \to 0 \qquad \text{a.s.} \tag{5.1}$$

then

$$\hat{\boldsymbol{\theta}}_n \to (\mathbf{a}_1,b_1) \qquad \text{a.s.} \quad [F]$$

Proof: The theorem follows from Theorem 1 of Huber, 1967 after we show that $\hat{\theta}_n$ almost surely ultimately stays in a compact set.

To show that $\hat{\boldsymbol{\theta}}_n$ almost surely ultimately stays in a compact set we argue as follows. By hypothesis, there exists $T>0$, $M>0$, $\delta>0$ and $\varepsilon>0$ such that

(a) $\rho(t)\geq M \quad \forall \ |t|\geq T$

and

(b) $M(1-\delta) > E_F\{\rho(\mathbf{a}'_1\mathbf{x}-b_1)\} + \varepsilon$.

Let $K_1>0$ be such that

$$P_F(\|\mathbf{x}\|\leq K_1) \geq 1-\delta/2$$

and let $K>0$ be such that $|b|>K$ and $\|\mathbf{x}\|\leq K_1$ implies that

$$|\mathbf{a}'\mathbf{x} - b| \geq T \qquad \forall\ ||\mathbf{a}|| = 1 .$$

Now, by the strong law of large numbers we have that for all $||\mathbf{a}|| = 1$ and $|b| > K$

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\mathbf{a}'\mathbf{x}_i - b) \geq \frac{1}{n}\sum_{i=1}^{n}\rho(\mathbf{a}'\mathbf{x}_i - b)I(||\mathbf{x}_i|| \leq K_\delta)$$

$$\geq M\frac{1}{n}\sum_{i=1}^{n}I(||\mathbf{x}_i|| \leq K_\delta) \geq M(1-\delta)$$

$$> E_F\{\rho(\mathbf{a}'_1\mathbf{x} - b_1)\} + \varepsilon \geq \frac{1}{n}\sum\rho(\mathbf{a}'_1\mathbf{x}_i - b_1) + \varepsilon/2$$

for sufficiently large n, a.s. $[F]$. Therefore, for sufficiently large n, $|b_n| \leq K$ and $\hat{\boldsymbol{\theta}}_n$ belongs to the compact set

$$[-K, K] \times \{\mathbf{a}: ||\mathbf{a}|| = 1\} \quad \blacksquare$$

Corollary 5.1: Assume that

(i) The loss function ρ is continuous, non-negative, and non-decreasing on $[0,\infty)$.

(ii) (a) The distribution of $\boldsymbol{\varepsilon}$ is spherically symmetric.

(b) The density f of $Y \equiv \mathbf{a}'\boldsymbol{\varepsilon}$ is unimodal and continuous.

(iii) $E\{\rho(\mathbf{a}'\boldsymbol{\varepsilon})\} < \infty$.

(iv) There exists $\delta > 0$ such that ρ and f are strictly monotone on $[0,\delta)$.

If $\hat{\boldsymbol{\theta}}_n$ satisfies (5.1) then

$$\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0 = (\mathbf{a}_0, b_0) \quad \text{a.s. } [F] \ \blacksquare \tag{5.2}$$

The proof of Corollary 5.1 uses the following lemma.

Lemma 5.1: Let, for each $t \in \mathbb{R}$

$$g(t) = E\{\rho(Y-t) - \rho(Y)\} \tag{5.3}$$

where Y is as defined in (ii.b) above. Then

$$g(t) > 0 \qquad \forall\ t \neq 0 \quad \blacksquare \tag{5.3'}$$

Proof: See the Appendix.

Proof of Corollary 5.1: The corollary follows from Theorem 5.1 once we show that our estimator is Fisher consistent over the distributions specified by (ii) and (iv), namely that

$$E_F\{\rho(\mathbf{a}'\mathbf{x}-b) - \rho(\mathbf{a}'_0\mathbf{x}-b_0)\} \geq 0 \qquad \forall\ \|\mathbf{a}\| = 1,\ \ b \in \mathbf{R} \tag{5.4}$$

with equality iff $(\mathbf{a}, b) = \boldsymbol{\theta}_0$ or $(\mathbf{a}, b) = -\boldsymbol{\theta}_0$.

Since (5.3) trivially holds when $E\{\rho(\mathbf{a}'\mathbf{x}-b)\} = \infty$ we only need to consider the case when $\rho(\mathbf{ax}-b)$ is integrable. In this case by (ii) and (iii) and Lemma 5.1

$$E\{\rho(\mathbf{a}'\mathbf{x}-b) - \rho(\mathbf{a}_0'\mathbf{x}-b_0)\} = E\{\rho(\mathbf{a}'\mathbf{x}-b) - \rho(\mathbf{a}'\boldsymbol{\varepsilon})\} =$$

$$= \int_{\mathbf{R}^{p+1}} \left[\int_{\mathbf{R}} [\rho(y - (b - \mathbf{a}'\mathbf{z})) - \rho(y)]dF_Y(y)\right] dF_{\mathbf{Z}}(\mathbf{z}) =$$

$$= \int_{\mathbf{R}^{p+1}} g(b - \mathbf{a}'\mathbf{z})dF_{\mathbf{Z}}(\mathbf{z}) \geq 0 . \tag{5.5}$$

Finally, if the last term of (5.5) is equal to 0 then $g(b - \mathbf{a}'\mathbf{X}) = 0$ a.s., which implies that $\mathbf{a}'\mathbf{X} - b = 0$ a.s. and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ or $\boldsymbol{\theta} = -\boldsymbol{\theta}_0$. $\blacksquare$

The assumption that $\mathbf{A}\boldsymbol{\varepsilon}$ is spherically symmetric is weaker than the assumption that ε_0 and ε_1 are independent, with $\lambda = \mathrm{Var}(\varepsilon_0)/\mathrm{Var}(\varepsilon_1)$ known, in the sense that the former does not require the existence of second moments.

On the other hand, the assumption that $A\varepsilon$ is spherically symmetric is a stronger assumption in the sense that it ensures the consistency result of Corollary 5.1. We have not been able to obtain this result with the "λ known" assumption.

Suppose that ε_0 and ε_1 are independent with distribution F and $\lambda = 1$. Then the distribution of $\boldsymbol{\varepsilon}$ is not spherically symmetric unless ε_0 and ε_1 are normal, as can be easily shown following the style of the proof of Theorem 4.6.4 in Chung (1974). If in this situation F is non-normal, then it appears that an ORM-estimator *will not* in general be consistent, while the classical OR-estimator *will* be consistent. However, numerical computations of the asymptotic bias under various non-spherically symmetric contaminated normal models (see Table 5.1 below) reveals, at least for the particular ORM-estimator under consideration, that (i) for $\lambda \neq 1$ the asymptotic bias is much smaller than that of the OR-estimate, and (ii) when $\lambda = 1$ $(\tau_1^2 = \tau_2^2)$ the computed values of the ORM are so close to zero that one suspects they are exactly equal to zero. However, we have not yet been able to prove this for any non-sphericaliy symmetric distribution. This point may deserve further attention.

In summary: the fact that an ORM-estimator may be asymptotically slightly biased in situations under which the OR-estimator is consistent is not so serious an objection as it might at first appear to be, since from the robustness point of view this small bias may be regarded as an "insurance fee" we are willing to pay to achieve a high degree of bias robustness and thereby avoid the catastrophic behavior of the OR-estimator in the presence of outliers.

Asymptotic Bias under Contaminated Gaussian EV Models

As we have seen above, ORM-estimators are asymptotically unbiased under a spherically symmetric E-V model. We may well ask what happens when the distribution of the error term of the E-V model is a non-spherically-symmetric outlier generating distribution. As an example we consider the very simple EV-contaminated model: $p = 1$, $\beta_1 = 1$, $\beta_0 = 0$, $X_0 \sim N(0,4)$, and $\varepsilon = \mathbf{e} + (B_1\Delta_1, B_2\Delta_2)$, where X_0, B_1, B_2, Δ_1, and Δ_2 are independent, $B_j \sim$ Binomial$(1,\gamma)$ and $\Delta_j \sim N(0,\tau_j^2)$, $(j = 1,2)$. In this case the classical orthogonal regression estimator, as well as the **ORM**-estimator, is asymptotically biased. Meanwhile, as illustrated in Table 5.1 below, the asymptotic bias of the **ORM**-estimator can be radically smaller than the asymptotic bias of the classical **OR** estimator. This speaks well for the bias robustness properties of the ORM-estimator. The particular ORM-estimator considered here is the one with (Tukey's) loss function

$$\rho_c(t) = \begin{cases} \dfrac{x^2}{2} - \dfrac{x^4}{2c^2} + \dfrac{x^6}{6c^4} & \text{for } |x| \leq c \\ \dfrac{c^2}{6} & \text{for } |x| \geq c \end{cases} \tag{5.5'}$$

with c = 4.7.

Of course further study is called for, and we intend to carry out analogous computations for higher order models, a richer class of contaminating distribution, and also some Monte Carlo simulations to obtain finite sample bias.

Table 5.1

		$\gamma=0.50$		$\gamma=0.10$		$\gamma=0.20$		$\gamma=0.25$	
τ_1	τ_2	OR	ORM	OR	ORM	OR	ORM	OR	ORM
0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	-0.01	-0.01	-0.01	-0.01	-0.02	-0.02	-0.03	-0.03
	2	-0.02	-0.02	-0.05	-0.03	-0.09	-0.06	-0.12	-0.08
	5	-0.14	-0.02	-0.26	-0.05	-0.45	-0.10	-0.51	-0.14
	10	-0.45	-0.01	-0.65	-0.03	-0.81	-0.07	-0.84	-0.10
1	0	0.01	0.01	0.01	0.01	0.03	0.02	0.03	0.03
	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	-0.02	-0.01	-0.04	-0.02	-0.07	-0.04	-0.09	-0.06
	5	-0.14	-0.02	-0.26	-0.03	-0.43	-0.08	-0.50	-0.12
	10	-0.44	-0.01	-0.65	-0.02	-0.81	-0.05	-0.84	-0.07
2	0	0.03	0.02	0.05	0.03	0.11	0.07	0.13	0.09
	1	0.02	0.01	0.04	0.02	0.08	0.04	0.10	0.06
	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	-0.12	-0.01	-0.23	-0.02	-0.40	-0.04	-0.46	-0.06
	10	-0.43	0.00	-0.64	0.00	-0.80	-0.01	-0.84	-0.02
5	0	0.17	0.02	0.36	0.05	0.80	0.12	1.05	0.16
	1	0.16	0.02	0.34	0.04	0.77	0.09	1.00	0.14
	2	0.14	0.01	0.30	0.02	0.65	0.04	0.85	0.07
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	-0.36	0.01	-0.57	0.01	-0.75	0.03	-0.80	0.04
10	0	0.80	0.01	1.85	0.03	4.19	0.08	5.41	0.10
	1	0.79	0.01	1.83	0.02	4.14	0.05	5.35	0.08
	2	0.77	0.00	1.76	0.00	4.00	0.01	5.16	0.02
	5	0.57	-0.01	1.31	-0.01	3.00	-0.03	3.89	-0.04
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

5.1.2 Asymptotic Normality

This subsection is devoted to the proof of the asymptotic normality of ORM-estimators. The assumptions under which asymptotic normality is proven are:

(1) $(\hat{\mathbf{a}}_n, \hat{b}_n) = \hat{\boldsymbol{\theta}}_n \to \theta_0 = (\mathbf{a}_0, b_0)$ a.s. as $n \to \infty$.

(2) ψ is symmetric, bounded and $\psi'(-t) = \psi'(t) \quad \forall \quad t \in \mathbb{R}$.

(3) $E\{\psi(\mathbf{a}'_0\mathbf{x}-b_0)\mathbf{x}\} = E\{\psi(\mathbf{a}'_0\mathbf{x}-b_0)\mathbf{a}'_0\mathbf{x}\}\cdot\mathbf{a}_0$.

(4) $E\{\psi(\mathbf{a}'_0\varepsilon)\mathbf{a}'_0\varepsilon\} > 0$.

(5) $E\{\|\mathbf{x}\|^3\} < \infty$.

NOTE: Assumption (3) holds whenever $(\mathbf{a}_0, b_0)$ minimizes $E\{\rho(\mathbf{a}'\mathbf{x}-b)\}$ among unitary vectors $\mathbf{a}$ in $\mathbb{R}^{p+1}$ and b in $\mathbb{R}$, for the Lagrangian of this optimization problem is

$$L(\mathbf{a},b,\lambda) = E\{\rho(\mathbf{a}'\mathbf{x}-b)\} - \frac{1}{2}\lambda(\mathbf{a}'\mathbf{a}-1) \tag{5.6}$$

and

$$\frac{\partial}{\partial \mathbf{a}} L = E\{\psi(\mathbf{a}'\mathbf{x}-b)\mathbf{x}\} - \lambda\mathbf{a} = 0 \tag{5.7}$$

$$\frac{\partial}{\partial b} L = E\{\psi(\mathbf{a}'\mathbf{x}-b)\} = 0 \ . \tag{5.8}$$

Premultiplying (5.6) by $\mathbf{a}'$ and using $\mathbf{a}'\mathbf{a} = 1$ we get

$$\lambda = E\{\psi(\mathbf{a}'\mathbf{x}-b)\mathbf{a}'\mathbf{x}\} \ . \tag{5.9}$$

To begin we show that ORM-estimators satisfy an equation which has a non-singular derivative.

First we consider models without an intercept term. Recall that in this case $\hat{\mathbf{a}}_n$ minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\rho(\mathbf{a}'\mathbf{x}_i) \tag{5.10}$$

among vectors $\mathbf{a}$ satisfying $\|\mathbf{a}\| = 1$. Equivalently, $\hat{\mathbf{a}}_n$ minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{\mathbf{a}'\mathbf{x}_i}{\|\mathbf{a}\|}\right) \tag{5.11}$$

among vectors $\mathbf{a}$ satisfying $\|\mathbf{a}\| \neq 0$. Now we differentiate (5.11) with respect to $\mathbf{a}$

to get the defining equation

$$(I - \hat{\mathbf{a}}_n \hat{\mathbf{a}}'_n)\, \frac{1}{n} \sum_{i=1}^{n} \psi(\hat{\mathbf{a}}'_n \mathbf{x}_i)\mathbf{y}_i \;=\; 0 \;. \tag{5.12}$$

REMARK: In fact

$$\frac{\partial}{\partial \mathbf{a}}\left[\rho\left(\frac{\mathbf{a}'\mathbf{x}_i}{\|\mathbf{a}\|}\right)\right] \;=\; \left(\mathrm{I} - \frac{\mathbf{a}\mathbf{a}'}{\|\mathbf{a}\|^2}\right) \frac{1}{n} \sum \psi\left(\frac{\mathbf{a}'\mathbf{x}_i}{\|\mathbf{a}\|}\right)\mathbf{x}_i \tag{5.13}$$

but we may choose our estimate to have length one, so that (5.12) holds true.

Let

$$\boldsymbol{\psi}(\mathbf{x}_i;\mathbf{a}) \;=\; (\mathrm{I} - \mathbf{a}\mathbf{a}')\psi(\mathbf{a}'\mathbf{x}_i)\mathbf{x}_i \;, \qquad \forall\, \mathbf{a} \in \mathbb{R}^p \tag{5.14}$$

and

$$\boldsymbol{\lambda}(\mathbf{a}) \;=\; E\{\boldsymbol{\psi}(\mathbf{x};\mathbf{a})\} \qquad \forall\, \mathbf{a} \in \mathbb{R}^p \;. \tag{5.15}$$

<u>*Lemma 5.3:*</u> Let

$$D\boldsymbol{\lambda}(\mathbf{a}) \;=\; \left(\frac{\partial \lambda_i}{\partial a_j}(\mathbf{a})\right)_{\substack{i=1\cdots p\\ j=1\cdots p}} \;. \tag{5.16}$$

Then $D\boldsymbol{\lambda}(\mathbf{a}_0)$ is a non-singular matrix.

<u>*Proof:*</u> It can be easily verified that

$$D\boldsymbol{\lambda}(\mathbf{a}) \;=\; E\{(\mathrm{I} - \mathbf{a}\mathbf{a}')\mathbf{x}\mathbf{x}'\psi'(\mathbf{a}'\mathbf{x})\} - E\{\psi(\mathbf{a}'\mathbf{x})[(\mathbf{a}'\mathbf{x})\mathrm{I} + \mathbf{a}\mathbf{x}']\} \tag{5.17}$$

and using assumption (3) we get

$$D\boldsymbol{\lambda}(\mathbf{a}_0) \;=\; (\mathrm{I} - \mathbf{a}_0\mathbf{a}'_0)\boldsymbol{\Sigma} - \alpha_0(\mathrm{I} + \mathbf{a}_0\mathbf{a}'_0) \tag{5.18}$$

where

$$\Sigma = E\{\psi'(\mathbf{a}'_0\mathbf{x})\mathbf{x}\mathbf{x}'\} = E\{\psi'(\mathbf{a}'_0\varepsilon)\varepsilon\varepsilon'\} + E\{\psi'(\mathbf{a}'_0\varepsilon)\}E\{\mathbf{x}\mathbf{x}'\} \tag{5.19}$$

and

$$\alpha_0 = E\{\psi(\mathbf{a}'_0\mathbf{x})\mathbf{a}'_0\mathbf{x}\} = E\{\psi(\mathbf{a}'_0\varepsilon)\mathbf{a}'_0\varepsilon\} . \tag{5.20}$$

Let $\mathbf{a} \in \mathbb{R}^p$. We will show that $[D\boldsymbol{\lambda}(\mathbf{a}_0)]\mathbf{a} = 0 \quad => \quad \mathbf{a} = 0$.

Case 1: $\mathbf{a} = \alpha\mathbf{a}_0$ for some $\alpha \in \mathbb{R}$

$$\begin{aligned} 0 &= \mathbf{a}'[D\boldsymbol{\lambda}(\mathbf{a}_0)] = \alpha\mathbf{a}'_0(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\Sigma - \alpha\alpha_0\mathbf{a}'_0(\mathbf{I} + \mathbf{a}_0\mathbf{a}_0) \\ &= -2\alpha\alpha_0\mathbf{a}_0 = 0 \quad => \quad \alpha = 0 \end{aligned} \tag{5.21}$$

since $\alpha_0 = E\{\psi(\mathbf{a}_0\varepsilon)\mathbf{a}'_0\varepsilon\} \neq 0$ by assumption (3).

Case 2: $\mathbf{a} \neq \alpha\mathbf{a}_0$ for any $\alpha \in \mathbb{R}$:

$$0 = \mathbf{a}'[D\boldsymbol{\lambda}(\mathbf{a}_0)] = \mathbf{a}(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\Sigma - \alpha_0\mathbf{a}'(\mathbf{I} + \mathbf{a}_0\mathbf{a}'_0) . \tag{5.22}$$

This implies that

$$0 = \mathbf{a}'[D\boldsymbol{\lambda}(\mathbf{a}_0)](\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\mathbf{a} = \mathbf{a}'(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)(\Sigma - \alpha_0\mathbf{I})(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\mathbf{a} \tag{5.23}$$

because $(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)(\mathbf{I} + \mathbf{a}_0\mathbf{a}'_0) = (\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)$ and $(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)^2 = (\mathbf{I} - \mathbf{a}_0\mathbf{a}_0)$. We notice that if $\mathbf{c} = (\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\mathbf{a}$ then $\mathbf{c}'\mathbf{a}_0 = \mathbf{a}'(\mathbf{I} - \mathbf{a}_0\mathbf{a}'_0)\mathbf{a}_0 = 0$.

The lemma follows after we prove the following claim.

Claim: If $\mathbf{c}'\mathbf{a}_0 = 0$ then $\mathbf{c}'\,\Sigma\,\mathbf{c} \neq \alpha_0\mathbf{c}'\mathbf{c}$ unless $\mathbf{c} = 0$.

Proof: $E\{\rho(\mathbf{a}'\varepsilon)\}$ is a continuous (in fact constant) function on the compact set $C_k = \{\mathbf{a} : \mathbf{a}'\mathbf{a} = k\}$. Because of Lagrange's multipliers theorem there exist $\tilde{\mathbf{a}}_0$ and β such that

$$\tilde{\mathbf{a}}'_0\tilde{\mathbf{a}}_0 = k \tag{5.24}$$

and

$$E\{\psi(\tilde{\mathbf{a}}'_0\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = \beta\tilde{\mathbf{a}}_0 \tag{5.25}$$

Hence

$$E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = \beta\mathbf{a} \qquad \forall\, \mathbf{a} \in C_k \ . \tag{5.26}$$

In fact, if P is an orthogonal matrix then

$$\tilde{\boldsymbol{\varepsilon}} = \mathrm{P}\boldsymbol{\varepsilon} \ , \quad \boldsymbol{\varepsilon} = \mathrm{P}'\tilde{\boldsymbol{\varepsilon}} \ , \quad \boldsymbol{\varepsilon} \cong \tilde{\boldsymbol{\varepsilon}}$$

$$\mathbf{a} = \mathrm{P}\tilde{\mathbf{a}}_0 \ , \quad \tilde{\mathbf{a}}_0 = \mathrm{P}'\mathbf{a} \tag{5.27}$$

and

$$\beta\tilde{\mathbf{a}}_0 = E\{\psi(\tilde{\mathbf{a}}_0'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = E\{\psi(\tilde{\mathbf{a}}'_0\mathrm{P}'\tilde{\boldsymbol{\varepsilon}})\mathrm{P}'\tilde{\boldsymbol{\varepsilon}}\} = \mathrm{P}'E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} \quad \Rightarrow$$

$$E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = \beta\mathrm{P}\tilde{\mathbf{a}}_0 = \beta\mathbf{a} \ . \tag{5.28}$$

It follows that

$$E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = \beta(|\mathbf{a}|)\mathbf{a} \qquad \forall\, \mathbf{a} \in \mathbb{R}^p \ . \tag{5.29}$$

Hence

$$\frac{\partial}{\partial \mathbf{a}} E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = \frac{\partial}{\partial \mathbf{a}}[\beta(\mathbf{a})\mathbf{a}] = \beta(\mathbf{a})I + \mathbf{a}\left[\frac{\partial}{\partial \mathbf{a}}\beta(\mathbf{a})\right]' \ . \tag{5.30}$$

On the other hand

$$\frac{\partial}{\partial \mathbf{a}} E\{\psi(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\} = E\{\psi'(\mathbf{a}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\} \ . \tag{5.31}$$

Thus

$$\Sigma = E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\} + E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})\}E\{\mathbf{x}\mathbf{x}'\} =$$

$$\alpha_0\mathrm{I} + \mathbf{a}_0\left[\frac{\partial}{\partial \mathbf{a}}\beta(\mathbf{a}_0)\right]' + E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon}\}E\{\mathbf{x}\mathbf{x}'\} \tag{5.32}$$

by (5.30), (5.31) and recalling that $\alpha_0 = \beta(\mathbf{a}_0)$.

Now suppose that

$$\mathbf{a}'\Sigma\mathbf{a} = \alpha_0\mathbf{a}'\mathbf{a} \ . \tag{5.33}$$

Then

$$\mathbf{a}'\,\Sigma\,\mathbf{a} = \alpha_0\mathbf{a}'\mathbf{a} + E\{\psi'(\mathbf{a}'_0\varepsilon)\}E\{(\mathbf{a}'\mathbf{x})^2\} \tag{5.34}$$

and this is impossible unless $\mathbf{a}'\mathbf{X} = 0$ a.s.

We now consider the general model. The estimators $(\hat{\mathbf{a}}_n, \hat{b}_n)$ satisfy the equations

$$\left.\begin{aligned} (\mathrm{I} - \hat{\mathbf{a}}_n\hat{\mathbf{a}}'_n)\,\frac{1}{n}\sum_{i=1}^{n}\psi(\hat{\mathbf{a}}'_n\mathbf{x}_i - \hat{b}_n)\mathbf{x}_i &= 0 \\ -\frac{1}{n}\sum_{i=1}^{n}\psi(\hat{\mathbf{a}}'_n\mathbf{x}_i - \hat{b}_n) &= 0 \end{aligned}\right| \tag{5.35}$$

or equivalently

$$\left.\begin{aligned} (\mathrm{I} - \hat{\mathbf{a}}_n\hat{\mathbf{a}}'_n)\,\frac{1}{n}\sum_{i=1}^{n}\psi(\hat{\mathbf{a}}'_n\mathbf{x}_i - \hat{b}_n)(\mathbf{x}_i - \mathbf{u}) &= 0 \\ -\frac{1}{n}\sum\psi(\hat{\mathbf{a}}'_n\mathbf{x}_i - \hat{b}_n) &= 0 \end{aligned}\right| \tag{5.36}$$

for any constant vector $\mathbf{u}$ and in particular for $\mathbf{u} = E\{\mathbf{x}\}$.

Let

$$\boldsymbol{\psi}(\mathbf{x};\mathbf{a},b) = \begin{bmatrix} (\mathrm{I} - \mathbf{a}\mathbf{a}')\psi(\mathbf{a}'\mathbf{x} - b)(\mathbf{x} - \mathbf{u}) \\ -\psi(\mathbf{a}'\mathbf{x} - b) \end{bmatrix} \tag{5.37}$$

and

$$\boldsymbol{\lambda}(\mathbf{a},b) = E\{\boldsymbol{\psi}(\mathbf{x};\mathbf{a},b)\} \ . \tag{5.38}$$

Now

$$D\boldsymbol{\lambda}(\mathbf{a}_0,b_0) = \begin{vmatrix} (\mathrm{I}-\mathbf{a}_0\mathbf{a}'_0)\Sigma - \alpha_0(\mathrm{I}+\mathbf{a}_0\mathbf{a}'_0) & (\mathrm{I}-\mathbf{a}_0\mathbf{a}'_0)E\{\psi'(\mathbf{a}'_0\varepsilon)(\varepsilon+\mathbf{x}-\mathbf{u})\} \\ (\mathrm{I}-\mathbf{a}_0\mathbf{a}'_0)E\{\psi'(\mathbf{a}'_0\varepsilon)(\varepsilon+\mathbf{x})\} & E\{\psi'(\mathbf{a}'_0\varepsilon)\} \end{vmatrix} \tag{5.39}$$

where

$$\Sigma = E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})(\mathbf{x}-\mathbf{u})(\mathbf{x}-\mathbf{u})'\} \tag{5.40}$$

and

$$\alpha_0 = E\{\psi(\mathbf{a}'_0\varepsilon)\mathbf{a}'_0\varepsilon\} \ . \tag{5.41}$$

We notice that by assumption (2)

$$\begin{aligned}(\mathbf{I}-\mathbf{a}_0\mathbf{a}'_0)E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})(\varepsilon+\mathbf{x}-\mathbf{u})\} &= (\mathbf{I}-\mathbf{a}_0\mathbf{a}'_0)[E\{\psi'(\mathbf{a}'_0\varepsilon)\boldsymbol{\varepsilon}\} + E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})\}E\{\mathbf{x}-\mathbf{u}\}] \\ &= \mathbf{0} \ . \end{aligned} \tag{5.42}$$

Finally

$$\det[D\boldsymbol{\lambda}(\mathbf{a}_0,b_0)] = E\{\psi'(\mathbf{a}'_0\boldsymbol{\varepsilon})\}\det[(\mathbf{I}-\mathbf{a}_0\mathbf{a}'_0)\Sigma - \alpha_0(\mathbf{I}+\mathbf{a}_0\mathbf{a}'_0)] \neq 0 \ . \tag{5.43}$$

Now we want to check the remaining conditions for asymptotic normality of Huber (1967). Let

$$\boldsymbol{\theta} = (\mathbf{a},b) \quad \text{and} \quad \boldsymbol{\theta}_0 = (\mathbf{a}_0,b_0) \ . \tag{5.44}$$

<u>*Lemma 5.4:*</u> There exist $\delta > 0$ and $\delta_0 > 0$ such that

$$||\boldsymbol{\lambda}(\boldsymbol{\theta})|| \geq \delta||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \quad \forall \ ||\theta-\boldsymbol{\theta}_0|| < \delta_0 \ . \tag{5.45}$$

Let

$$\boldsymbol{\psi}(\mathbf{x};\boldsymbol{\theta}) = (\mathbf{I}-\mathbf{aa})\psi(\mathbf{a}'\mathbf{x}-b)\mathbf{x} \tag{5.46}$$

and

$$u(\mathbf{x};\boldsymbol{\theta},d) = \sup_{||\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}||\leq d} ||\boldsymbol{\psi}(\mathbf{x};\boldsymbol{\theta}) - \boldsymbol{\psi}(\mathbf{x};\tilde{\boldsymbol{\theta}})|| \ . \tag{5.47}$$

Proof: See the Appendix.

Lemma 5.5: There exist $\beta_1 > 0$, $\beta_2 > 0$, and $\delta_0 > 0$ such that

$$E\{u(\mathbf{x};\boldsymbol{\theta},d)\} \leq \beta_1 d \qquad \text{for } \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq \delta_0 \tag{5.48}$$

and

$$E\{u^2(\mathbf{x};\theta,d)\} \leq \beta_2 d \qquad \text{for } \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + d \leq \delta_0 \ . \tag{5.49}$$

Proof: See the Appendix.

Lemmas 5.3, 5.4 and 5.5 show that all the conditions for the Corollary to Theorem 3 of Huber (1967) hold. In our case we have

$$\mathbf{\Lambda} = D(\boldsymbol{\lambda}(\mathbf{a}_0, b_0)) \tag{5.50}$$

and

$$\mathbf{C} = cov[\boldsymbol{\psi}(\mathbf{x};\mathbf{a}_0,b_0)] \tag{5.51}$$

where $\boldsymbol{\psi}(\mathbf{x};\mathbf{a}_0,b_0)$ is defined in (5.37). Thus, we have established the following Theorem.

Theorem 5.2: If assumptions (1)-(5) hold then

$$\sqrt{n}\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{V}) \tag{5.52}$$

where $\mathbf{V} = \mathbf{\Lambda}^{-1}\mathbf{C}(\mathbf{\Lambda}')^{-1}$.

5.1.3 Qualitative Robustness of the ORM-Estimator

Theorem 5.3: If ρ is bounded and non-decreasing on $[0,\infty)$ then the ORM-estimator is *qualitatively robust* at any distribution F for which $E_F\{\rho(\mathbf{a}'\mathbf{x}-b)\}$ has a unique minimum.

Proof: To prove qualitative robustness of the **ORM**-estimator it is enough to show that this estimator is weakly continuous when viewed as a functional on the set of distribution functions (cf. Huber, 1981). Let $F^{(m)}$ be a sequence of distribution functions on $\mathbb{R}^{p+1}$ which converges weakly toward F. We will show that $\boldsymbol{\theta}_m = \theta(F^{(m)})$ can be uniquely defined for $m \geq N$ and that $\boldsymbol{\theta}_m \to \boldsymbol{\theta}_0 = \boldsymbol{\theta}(F)$ as $m \to \infty$. First we notice that the family of functions

$$\{\rho(\mathbf{x}'\mathbf{a}-b): \ \|\mathbf{a}\|=1, \ b \in \mathbb{R}\} \tag{5.53}$$

is uniformly bounded and equicontinuous at all $\mathbf{x} \in \mathbb{R}^{p+1}$ (i.e., $\forall$ $\mathbf{x}\in\mathbb{R}^{p+1}$, $\forall$ $\varepsilon > 0$, there exists $\delta > 0$ with the property that $\|\mathbf{x}-\mathbf{y}\| < \delta$ implies $|\rho(\mathbf{a}'\mathbf{x}-b)-\rho(\mathbf{a}'\mathbf{y}-b)| < \varepsilon$, $\forall$ $\|\mathbf{a}\| = 1$ and $\forall$ $b \in \mathbb{R}$).

Therefore

$$\tau_m(\mathbf{a},b) \equiv \int_{\mathbb{R}^{p+1}} \rho(\mathbf{a}'\mathbf{x}-b)dF^{(m)}(\mathbf{x}) \to \int_{\mathbb{R}^{p+1}} \rho(\mathbf{a}'\mathbf{x}-b)dF(\mathbf{x}) \equiv \tau(\mathbf{a},b) \tag{5.54}$$

uniformly on $\|\mathbf{a}\| = 1$, $b \in \mathbf{R}$ (see Billingsley, 1968, p. 17).

Let $\boldsymbol{\theta}_0 = (\mathbf{a}_0, b_0)$ be the point at which $\tau(\mathbf{a},b)$ achieves its unique minimum $\tau_0 = \tau(\mathbf{a}_0, b_0)$ and $M = \lim_{t\to\infty} \rho(t)$. Hence

$$\delta = M - \tau_0 > 0 . \tag{5.55}$$

Let $N \in \mathbb{N}$ and $K > 0$ be such that for all $m \geq N$, $|b| \geq K$, and $\|\mathbf{a}\| = 1$ we have

$$\tau_m(\mathbf{a},b) \geq \tau(\mathbf{a},b) - \frac{\delta}{4} \tag{5.56}$$

and

$$\tau(\mathbf{a}, b) \;\geq\; M - \frac{\delta}{4} \tag{5.57}$$

Thus, for all $m \geq N$, $|b| > K$, and $\|\mathbf{a}\| = 1$

$$\tau_m(\mathbf{a}, b) \;\geq\; \tau(\mathbf{a}, b) - \frac{\delta}{4} \;\geq\; M - \frac{\delta}{2} \;=\; \tau_0 + \frac{\delta}{2} \;. \tag{5.58}$$

On the other hand, for $m \geq N$

$$\tau_m(\mathbf{a}_0, b_0) \;\leq\; \tau(\mathbf{a}_0, b_0) + \frac{\delta}{4} \;=\; \tau_0 + \frac{\delta}{4} \;. \tag{5.59}$$

From (5.58) and (5.59) it follows that for all $m \geq N$, $\tau_m(\mathbf{a}, b)$ achieves its minimum on the compact set $\{(\mathbf{a}, b) : \|\mathbf{a}\| = 1\,,\ |b| \leq K\}$. We can uniquely define $\boldsymbol{\theta}_m = \boldsymbol{\theta}(F^{(m)})$ as in section 3.

In order to prove convergence of $\boldsymbol{\theta}_m$ to θ_0 let $\varepsilon > 0$ be fixed. A compactness argument similar to the one used by Huber, 1967 in the proof of Theorem 1 shows that there exists $\delta_0 > 0$ such that

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \;=\; \|\mathbf{a} - \mathbf{a}_0\| + |b - b_0| \;>\; \varepsilon \quad \Rightarrow \quad \tau(\boldsymbol{\theta}) \;>\; \tau(\boldsymbol{\theta}_0) + \delta_0 \;. \tag{5.60}$$

Therefore, there exists $N \in \mathbb{N}$ such that for all $m \geq N$

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \varepsilon \quad \Rightarrow \quad \tau_m(\boldsymbol{\theta}) > \tau(\boldsymbol{\theta}_0) + \delta_0 - \frac{\delta}{4} = \tau_0 + \frac{3}{4}\,\delta_0 \tag{5.61}$$

and the theorem follows because for all $m > N$

$$\tau_m(\boldsymbol{\theta}_0) \;\leq\; \tau(\boldsymbol{\theta}_0) + \frac{\delta_0}{4} \;. \tag{5.62}$$

5.2 FUNCTIONAL CASE

5.2.1 Consistency

Theorem 5.4: In addition to (5.2), (i), (ii) and (iv) of Theorem 5.1 assume that

(viii) $\lim_{t \to \infty} \rho(t) = M < \infty$.

(ix) For all $\varepsilon > 0$, $\exists$. > 0 such that

$$\lim_{n \to \infty} \frac{n(t)}{n} \geq 1 - \varepsilon \tag{5.63}$$

where

$$n(t) = \# J_n(t) \equiv \# \{ i \leq n : ||\mathbf{X}_i|| \leq t \} . \tag{5.64}$$

(x) For all $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\exists$ $\delta > 0$, $\varepsilon > 0$, and $N \in \mathbf{N}$ such that

$$\frac{m(\varepsilon)}{n} \geq \delta \ , \quad \forall \ n \geq N \tag{5.65}$$

where

$$m(\varepsilon) = \# I_n(\varepsilon) = \{ i : | \mathbf{a}'\mathbf{X}_i - b | \geq \varepsilon \} . \tag{5.66}$$

Then

$$\widehat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0 . \tag{5.67}$$

Furthermore, almost sure convergence of $\boldsymbol{\theta}_n$ to $\boldsymbol{\theta}_0$ holds provided that

(xi)

$$\frac{1}{n} \sum_{i=1}^{n} E \{ \rho(\mathbf{a}'\mathbf{x}_i - b) \} \to g(\mathbf{a}, b) \qquad \forall \ ||\mathbf{a}|| = 1 , \quad b \in \mathbb{R} \tag{5.68}$$

for some continuous function g.

NOTE: Assumption (ix) is equivalent to requiring that the sequence of empirical distributions

$$\mathbf{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{[0,t]}(\|\mathbf{X}_i\|) \tag{5.69}$$

is tight. Assumption (x) essentially means that besides $H(\mathbf{a}_0, b_0)$, there is no other hyperplane containing most of the $\mathbf{X}_i$. A sufficient condition for (ix) is that the largest eigenvalue λ_{pn} of

$$\frac{1}{n}\sum^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})' = \mathbf{S}_n \tag{5.70}$$

is bounded, while a sufficient condition for (x) is that the smallest eigenvalue $\lambda_{1n}(K)$

$$\frac{1}{n}\sum(\mathbf{X}_i - (\overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}}) I(\|\mathbf{X}_i\| \leq K) = \mathbf{S}_n(K) \tag{5.71}$$

is bounded away from 0 for some $K \geq 0$ and all $n \geq N > 0$.

The following lemmas are needed to prove Theorem 5.4. For proofs of these lemmas see the Appendix.

Lemma 5.6:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}\right] \to 0 \qquad \text{a.s.} \tag{5.72}$$

uniformly on $\|\mathbf{a}\| = 1$ and b.

Now Lemmas 5.1 and 5.6 are used to prove the following lemma.

Lemma 5.7: $\hat{\boldsymbol{\theta}}_n = (\hat{\mathbf{a}}_n, \hat{b}_n)$ is a.s. ultimately in a compact set.

Lemma 5.8: The family of functions

$$= \left\{ h_n(\mathbf{a}, b) : h_n(\mathbf{a}, b) \; = \; \frac{1}{n} \sum E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\} \right\} \tag{5.73}$$

is equicontinuous and pointwise bounded on

$$\mathbf{C} \; = \; \{\boldsymbol{\theta} : \boldsymbol{\theta} = (\mathbf{a}, b), \|\mathbf{a}\| = 1, b \in \mathbf{R}\} \; . \tag{5.74}$$

Lemma 5.9: For all $\boldsymbol{\theta} \in \mathbf{C}$, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ there exists $\Delta_{\boldsymbol{\theta}} > 0$ and $N \in \mathbf{N}$ such that

$$h_n(\boldsymbol{\theta}) \; \geq \; h_n(\boldsymbol{\theta}_0) + \Delta_{\boldsymbol{\theta}} \qquad \forall \; n \geq N \; . \tag{5.75}$$

Proof of Theorem 5.4: Lemma 5.6 is just a technicality needed, together with Lemma 5.1, to prove Lemma 5.7. The first part of the theorem is proved by showing that given a subsequence $\widehat{\boldsymbol{\theta}}_{n'}$, there exists a further subsequence $\widehat{\boldsymbol{\theta}}_{n''}$ such that

$$\widehat{\boldsymbol{\theta}}_{n''} \; \to \; \boldsymbol{\theta}_0 \qquad \text{a.s.} \; . \tag{5.76}$$

By the Arzela-Ascoli theorem, which can be applied to $\{h_n\}$ in view of Lemma 5.8, given a subsequence $h_{n'}(\boldsymbol{\theta})$ there exists a further subsequence $h_{n''}(\boldsymbol{\theta})$ and a continuous function $h(\boldsymbol{\theta})$ such that

$$\lim_{n'' \to \infty} h_{n''}(\boldsymbol{\theta}) \; = \; h(\boldsymbol{\theta}) \tag{5.77}$$

uniformly on $\mathbf{C}$.

Now Lemmas 5.7 and 5.9 allow us to prove the theorem by following the same method used by Huber (1967) to prove Theorem 2. The almost sure part of the theorem follows in the same way except that now assumption (ix) makes the subsequence argument no longer needed.

5.2.2 Asymptotic Normality

The proof of Theorem 5.2 was based on Huber (1967), which deals only with the case of i.i.d. observations. Under the functional model the observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are not i.i.d. because $E\{\mathbf{x}_i\} = \mathbf{X}_i$ with the $\mathbf{X}_i$ fixed and, in general, $\mathbf{X}_i \neq \mathbf{X}_j$ for $i \neq j$. Hence adapting the proof of Theorem 5.2 to cover the functional case could be done, extending Huber's 1967 to the independent, but *not* identically distributed, case. Such an extension is of interest in itself, but we won't pursue the issue here.

A Taylor expansion of the ORM estimating equation (2.18) gives (we only consider the no-intercept case here)

$$0 = \frac{1}{\sqrt{n}} \sum^{n} \psi[e_i(\hat{\boldsymbol\beta})]\mathbf{Z}_i(\hat{\boldsymbol\beta}) = \frac{1}{\sqrt{n}} \sum^{n} \psi[e_i(\boldsymbol\beta_0)]\mathbf{Z}_i(\boldsymbol\beta_0)$$

$$+ \frac{1}{n}\left\{\sum^{n} \frac{1}{1+\boldsymbol\beta'_0\boldsymbol\beta_0} \psi'[e_i(\boldsymbol\beta_0)](\mathbf{Z}_i(\boldsymbol\beta_0)\mathbf{Z}'_i(\boldsymbol\beta_0))\right.$$

$$\left. + \psi[e_i(\boldsymbol\beta_0)\frac{\partial}{\partial\boldsymbol\beta}\mathbf{Z}_i(\boldsymbol\beta_0)\right\} \sqrt{n}\,(\hat{\boldsymbol\beta} - \boldsymbol\beta_0) + R_n \tag{5.78}$$

Hence, under enough regularity conditions on the sequence $\mathbf{X}_1, \mathbf{X}_2, \cdots$, the distribution of ε and the score function ψ, $\sqrt{n}\,(\hat{\boldsymbol\beta} - \boldsymbol\beta_0)$ will be asymptotically equivalent to

$$\mathbf{A}^{-1}(\boldsymbol\beta_0)\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi[e_i(\boldsymbol\beta_0)]\mathbf{Z}_i(\boldsymbol\beta_0)$$

where

$$\mathbf{A} = \mathbf{Z}(\boldsymbol\beta_0) = p\lim\left\{\frac{1}{n} \sum \frac{1}{1+\boldsymbol\beta'_0\boldsymbol\beta_0} \psi'[e_i(\boldsymbol\beta_0)](\mathbf{Z}_i(\boldsymbol\beta_0)\mathbf{Z}'_i(\boldsymbol\beta_0))\right.$$

$$\left. + \frac{1}{n} \sum \psi[e_i(\boldsymbol\beta_0)] \frac{\partial}{\partial\boldsymbol\beta} \mathbf{Z}_i(\boldsymbol\beta_0)\right\} \tag{5.79}$$

Therefore, under regularity conditions we will have that

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to_d N\,[\,\mathbf{0},\, \mathbf{A}^{-1}\mathbf{V}(A^{-1})'\,] \tag{5.80}$$

where

$$\mathbf{V} = Cov(\psi[\mathbf{e}(\boldsymbol{\beta}_0)]\,\mathbf{Z}(\boldsymbol{\beta}_0)) \tag{5.81}$$

REMARK (qualitative robustness of ORM-estimators under the functional model): Theorem 5.2 of Huber, 1981 only applies to estimators which depend on i.i.d. observations. Therefore our proof of qualitative robustness for the ORM-estimators under the structural model will not carry over in a straightforward way to the functional model. The basic difficulty is that under the functional model the distribution of the stochastic process $\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots,$ cannot be characterized simply by a marginal distribution. The distributions F_i of the $\mathbf{x}_i$ have different locations $\mathbf{X}_i$. Thus the overall measure μ for the process $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ will depend upon the locations $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots,$ as well as on the common marginal distribution for the ε_i. Nevertheless our conjecture is that under appropriate regularity conditions, ORM-estimators with a bounded loss function are qualitatively robust. It should be possible to prove this via the approach used in the proof of Theorem 5.3, along with some reformulation of the asymptotic "estimating minimization problem" in the spirit of Theorem 5.4.

Appendix

MLE MOTIVATION (general p case):

To extend the MLE motivation of Section 2 to the general p case it is enough to notice that the system of linear equations (on $X_1, \ldots, X_p$)

$$
\begin{aligned}
(x_0-\beta_0-\beta_1X_1-\beta_2X_2-\cdots-\beta_pX_p)\beta_1+(x_1-X_1) &= 0\\
(x_0-\beta_0-\beta_1X_1-\beta_2X_2-\cdots-\beta_pX_p)\beta_2+(x_2-X_2) &= 0\\
&\vdots\\
(x_0-\beta_0-\beta_1X_1-\beta_2X_2-\cdots-\beta_pX_p)\beta_p+(x_p-X_p) &= 0
\end{aligned}
$$

has the (unique) solution

$$\hat{X}_k = x_k + \delta\beta_k e = \delta z_k \quad , \qquad k = 1, \ldots, p \, .$$

Hence, for $i = 1, \ldots, n$ we have that

$$(x_{0i}-\beta_0-\beta_1X_{1i}-\cdots-\beta_pX_{pi})^2+\sum_{k=1}^{p}(x_{ki}-X_{ki})^2 = e_i^2$$

and

$$(y_i-\beta_0-\beta_1X_{1i}-\cdots-\beta_pX_{pi})^2 = \delta^2 e_i^2 \, .$$

Therefore, the ORM-estimator with loss function $\tilde{\rho}(t) = \frac{1}{2}\rho(t)$ (cf. Section 2) has the same estimating equation that is obtained by differentiating the log-likelihood function

$$l(\boldsymbol{\beta}, \mathbf{X}_1, \ldots, \mathbf{X}_n ; \mathbf{x}_1, \ldots, \mathbf{x}_n) =$$

$$= n \log K - \sum^{n} \rho\left[(x_{0i}-\beta_0-\beta_1X_{1i}-\cdots-\beta_pX_{pi})^2+\sum_{k=1}^{p}(x_{ki}-X_{ki})^2\right] .$$

Proof of Lemma 2.1:

Let $\varepsilon > 0$ and $K = \max_{1 \le i \le n} \|\mathbf{x}_i\|$. Then for $i = 1, \ldots, n$

$$\|\mathbf{a}'\mathbf{x}_i - \mathbf{c}'\mathbf{x}_i\| \le \|\mathbf{a}-\mathbf{c}\| \cdot \|\mathbf{x}_i\| \le K\|\mathbf{a}-\mathbf{c}\| . \tag{A.1}$$

If $\|\mathbf{a}-\mathbf{c}\| < \varepsilon/K \equiv \delta$, then for $i = 1, \ldots, n$

$$\mathbf{c}'\mathbf{x}_i - \varepsilon \le \mathbf{a}'\mathbf{x}_i \le \mathbf{c}'\mathbf{x}_i + \varepsilon . \tag{A.2}$$

By monotonicity of ψ, it follows that for all $b_1 \in \mathrm{B}(\mathbf{a})$

$$\sum_{i=1}^{n} \psi[\mathbf{c}'\mathbf{x}_i - (b_1+\varepsilon)] \le \sum_{i=1}^{n} \psi[\mathbf{a}'\mathbf{x}_i - b_1] = 0 \le \sum_{i=1}^{n} \psi[\mathbf{c}'\mathbf{x}_i - (b_1-\varepsilon)] \tag{A.3}$$

and since $\sum_{i=1}^{n} \psi(\mathbf{c}'\mathbf{x}_i - b_2) = 0$, it follows that $b_1 - \varepsilon \le b_2 \le b_1 + \varepsilon$.

Proof of Lemma 2.2:

Let $\varepsilon > 0$, let δ be defined as in the proof of Lemma 2.1, and let $\|\mathbf{a}-\mathbf{c}\| < \delta$. We may assume without losing generality that $F(\mathbf{a}) \le F(\mathbf{c})$. By Lemma 2.1, if $b_0 = F(\mathbf{a})$ then there exists $b \in \mathrm{B}(\mathbf{c})$ such that $0 \le b - b_0 < \varepsilon$, and the lemma follows since

$$b \le b_0 + \varepsilon \quad \Rightarrow \quad F(\mathbf{c}) \le b_0 + \varepsilon \quad \Rightarrow \quad 0 \le F(\mathbf{c}) - F(\mathbf{a}) \le \varepsilon . \tag{A.4}$$

Proof of Lemma 5.1:

Let $t > 0$. By (i) and (iii-b)

$$g(t) = \int_{-\infty}^{\infty} [\rho(y-t) - \rho(y)] f(y) dy =$$

$$= \int_{-\infty}^{t/2} [\rho(y-t) - \rho(y)] f(y) dy + \int_{t/2}^{\infty} [\rho(y-t) - \rho(y)] f(y) dy =$$

$$= \int_{t/2}^{\infty} [\rho(-y) - \rho(t-y)] f(t-y) dy - \int_{t/2}^{\infty} [\rho(y) - \rho(y-t)] f(y) dy =$$

$$= \int_{t/2}^{\infty} [\rho(y) - \rho(y-t)][f(y-t) - f(y)] dy \geq 0 . \tag{A.5}$$

Finally,

$$[\rho(y) - \rho(y-t)][f(y-t) - f(y)]\Big|_{y=t} = [\rho(t) - \rho(0)][f(0) - f(t)] > 0 \tag{A.6}$$

and the claim follows by (iv) and continuity of ρ and f.

Proof of Lemma 5.4:

By definition of derivative we have that

$$||\lambda(\theta) - \lambda(\theta_0) - D\lambda(\theta_0)(\theta - \theta_0)|| = 0(||\theta - \theta_0||) \tag{A.7}$$

that is,

$$||\lambda(\theta) - D\lambda(\theta_0)(\theta - \theta_0)|| = 0(||\theta - \theta_0||) . \tag{A.8}$$

Given $\varepsilon > 0$, $\exists$ $\delta_0 > 0$ such that $|\theta - \theta_0| < \delta_0$ implies

$$||\lambda(\theta) - D\lambda(\theta_0)(\theta - \theta_0)|| < \varepsilon||\theta - \theta_0|| . \tag{A.9}$$

Now

$$\Big| \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| - \|D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\| \Big| \leq \|\boldsymbol{\lambda}(\boldsymbol{\theta}) - D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\| \leq \varepsilon\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|$$

$$=> \quad -\varepsilon\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| - \|D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\|$$

$$=> \quad \|\boldsymbol{\lambda}(\boldsymbol{\theta})\| \geq \|D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\| - \varepsilon\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \geq \delta\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \tag{A.10}$$

because

$$\|D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\| = \{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'[D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)]'[D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)](\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}^{\frac{1}{2}}$$

$$= \{(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'A(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}^{\frac{1}{2}} = d_A(\boldsymbol{\theta},\boldsymbol{\theta}_0) \tag{A.11}$$

where $A = [D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)]'[D\boldsymbol{\lambda}(\boldsymbol{\theta}_0)]$ is a positive definite matrix and $d_A(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ is a distance on $\mathbb{R}^{p+1}$; which is equivalent to the Euclidean distance on $\mathbb{R}^{p+1}$ (that is, there exist $\beta_1, \beta_2 > 0$ such that $\beta_1\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq d_A(\boldsymbol{\theta},\boldsymbol{\theta}_0) \leq \beta_2\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|$). We can choose $\varepsilon < \beta_1$, hence $\delta = \beta_1 - \varepsilon > 0$.

Proof of Lemma 5.5:

The lemma follows after we show that there exist $K_1 > 0$ and $K_2 > 0$ such that for all $\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}$

$$\|\boldsymbol{\psi}(\mathbf{x},\boldsymbol{\theta}) - \boldsymbol{\psi}(\mathbf{x},\tilde{\boldsymbol{\theta}})\| \leq K_1\|\mathbf{x}\|^2\,\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\| \tag{A.12}$$

and

$$\|\boldsymbol{\psi}(\mathbf{x},\boldsymbol{\theta}) - \boldsymbol{\psi}(\mathbf{x},\tilde{\boldsymbol{\theta}})\| \leq K_2\|\mathbf{x}\| . \tag{A.13}$$

In fact by assumption (2) there exist $K_1 > 0$ and $K_2 > 0$ such that for all $t_1, t_2 \in \mathbf{R}$

$$|\psi(t_1) - \psi(t_2)| \leq K_1|t_1 - t_2| \tag{A.14}$$

and

$$|\psi(t_1) - \psi(t_2)| \leq K_2 . \tag{A.15}$$

Let $\boldsymbol{\theta} = (\mathbf{a}, b)$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\mathbf{a}}, \tilde{b})$. Now

$$|(\mathbf{a}'\mathbf{x} - b) - (\tilde{\mathbf{a}}'\mathbf{x} - \tilde{b})| \le \|\mathbf{a} - \tilde{\mathbf{a}}\| \, \|x\| + |b - \tilde{b}| \le (1 + \|\mathbf{x}\|) \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\| \quad \text{(A.16)}$$

so that

$$\|\boldsymbol{\psi}(\mathbf{x}, \boldsymbol{\theta}) - \boldsymbol{\psi}(\mathbf{x}, \tilde{\boldsymbol{\theta}})\| = \| [\psi(\mathbf{a}'\mathbf{x} - b) - \psi(\tilde{\mathbf{a}}'\mathbf{x} - \tilde{b})] \mathbf{x} \| \le K_1 (1 + \|\mathbf{x}\|) \cdot \|\mathbf{x}\| \cdot \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\| \quad \text{(A.17)}$$

and

$$\|\boldsymbol{\psi}(\mathbf{x}, \boldsymbol{\theta}) - \boldsymbol{\psi}(\mathbf{x}, \tilde{\boldsymbol{\theta}})\| \le |\psi(\mathbf{a}'\mathbf{x} - b) - \psi(\tilde{\mathbf{a}}'\mathbf{x} - \tilde{b})| \cdot \|\mathbf{x}\| \le K_2 \|\mathbf{x}\| . \quad \text{(A.18)}$$

Proof of Lemma 5.6:

For each fixed $(\mathbf{a}, b)$ the lemma easily follows. Indeed, if $y_i = \rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}$ then $E\{y_i\} = 0$, $(1/n) \sum^{\infty} E\{y_i^2 / i^2\} \le 2M^2 \sum^{n} i^2 < \infty$. Hence Kolmsgorov's theorem implies that $(1/n) \sum^{n} y_i \to 0$, a.s. as $n \to \infty$.

Let $\varepsilon > 0$ be fixed and let $t \geq 0$:

$$\frac{1}{n} \sum^{n} [\rho(\mathbf{a}\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}] =$$

$$= \frac{1}{n} \sum^{n} [\rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}][I(\|\mathbf{X}_i\| > t) + I(\|\boldsymbol{\varepsilon}_i\| > t)]$$

$$+ \frac{1}{n} \sum^{n} [\rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}] I(\|\mathbf{X}_i\| \le t) I(\|\boldsymbol{\varepsilon}_i\| \le t) . \quad \text{(A.19)}$$

The first term on the right hand side of (A.19) is bounded by $2M[(1/n) \sum^{n} I(\|\mathbf{X}_i\| > t) + (1/n) \sum^{n} I(\|\boldsymbol{\varepsilon}_i\| > t)]$. By (ix) this can be done less than $\varepsilon/2$ for some $t > 0$ and all $n \geq N$.

The second term on the right hand side of (A.19) converges to 0 uniformly on $\mathbf{C}_k = \{(\mathbf{a}, b) : \|\mathbf{a}\| = 1, |b| \le K\}$. Indeed, if $(\mathbf{a}_1, b_1)$ and $(\mathbf{a}_2, b_2)$ are in C_k, for $i = 1, 2, \cdots$, we have

$$[\rho(\mathbf{a'}_1\mathbf{x}_i - b_1) - E\{\rho(\mathbf{a'}_1\mathbf{x}_i - b_1)\} - \rho(\mathbf{a'}_2\mathbf{x}_i - b_2) + $$
$$+ E\{\rho(\mathbf{a'}_2\mathbf{x}_i - b_2)\}]I(\|\mathbf{X}_i\| \le t)I(\|\boldsymbol{\varepsilon}_i\| \le t)$$
$$\le |\rho(\mathbf{a'}_1\mathbf{x}_i - b_1) - \rho(\mathbf{a'}_2\mathbf{x}_i - b_2)|I(\|\mathbf{X}_i\| \le t)I(\|\boldsymbol{\varepsilon}_i\| \le t) +$$
$$+ E\{|\rho(\mathbf{a'}_1\mathbf{x}_i - b_1) - \rho(\mathbf{a'}_2\mathbf{x}_i - b_2)|I(\|\mathbf{X}_i\| \le t)\} \ . \tag{A.20}$$

The first term on the right hand side of (A.20) can be done less than $\varepsilon/4$ for all $i \in \mathbb{N}$ and all $\|\mathbf{a}_1 - \mathbf{a}_2\| + |b_1 - b_2| < \delta$ for some $\delta > 0$. The same thing is true with respect to the second term on the right hand side of (A.20). (Divide the domain of integration into two parts, one of them being the set $\{\boldsymbol{\varepsilon}: \|\boldsymbol{\varepsilon}\| \le t_0\}$, for some t_0 large enough, and notice that for $\|\boldsymbol{\varepsilon}\| \le t_0$ and $\|\mathbf{X}_i\| \le t$, $\rho(\mathbf{a'}\mathbf{x}_i - b)$ is a uniformly continuous function of $(\mathbf{a}, b)$.) Now use pointwise convergence of the second term on the right hand side of (A.19), compactness of $\mathbf{C}_k$, and the just-proved continuity of $\rho(\mathbf{a'}\mathbf{x}_i - b) - E\{\rho(\mathbf{a'}\mathbf{x}_i - b)\}$ uniformly on $i \in \mathbf{N}$ and $(\mathbf{a}, b) \in \mathbf{C}_k$ to get the desired result.

Finally, $K > 0$ can be chosen so large that for all $|b| \ge K$, $\|\mathbf{a}\| = 1$, and $i \in \mathbb{N}$

$$M - \frac{\varepsilon}{4} \le \rho(\mathbf{a'}\mathbf{x}_i - b)I(\|\boldsymbol{\varepsilon}_i\| \le t)I(\|\mathbf{X}_i\| \le t) \le M \tag{A.21}$$

and

$$M - \frac{\varepsilon}{4} \le E\{\rho(\mathbf{a'}\mathbf{x}_i - b)I(\|\mathbf{X}_i\| \le t)\} \le M \ . \tag{A.22}$$

Hence, the second term on the right hand side of (A.19) can be done less than $\varepsilon/2$ for all $n \ge N$, $\|\mathbf{a}\| = 1$ and $b \in \mathbf{R}$, proving the lemma.

Proof of Lemma 5.7:

Since $\|\hat{\mathbf{a}}_n\| = 1$, we only need to show that there exists $K > 0$ such that $|\hat{b}_n| \leq K$ a.s. for large n.

By Lemma 5.1 there exist $0 < p < 1$, $t_0 > 0$, and $\varepsilon > 0$ such that for all $|t| \geq t_0$

$$pE\{\rho(\mathbf{a}'\boldsymbol{\varepsilon} + t)\} > E\{\rho(\mathbf{a}'\boldsymbol{\varepsilon})\} + \varepsilon \,. \tag{A.23}$$

By (ix) there exist $N \in \mathbb{N}$ and $t_1 > 0$ such that for all $n \geq N$ and $|t| \geq t_1$

$$\frac{n(t)}{n} \geq p \,. \tag{A.24}$$

Let $t_2 = \max\{t_0, t_1\}$. If $\mathbf{X}_i \in J_n(t_2)$ and $|b| > 2t_2$ then

$$|b - \mathbf{a}'\mathbf{X}_i| \geq |b| - |\mathbf{a}'\mathbf{X}_i| \geq |b| - \|\mathbf{X}_i\| \geq |b| - t_2 > t_0 \,. \tag{A.25}$$

Finally, for $|b| > 2t_2$

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{a}'\mathbf{x}_i - b) = \frac{1}{n}\sum_{i=1}^{n} E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\} + \frac{1}{n}\sum_{i=1}^{n}[\rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}]$$

$$\geq \frac{1}{n}\sum_{i \in J_n(t_2)} E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\} + \frac{1}{n}\sum_{i=1}^{n}[\rho(\mathbf{a}'\mathbf{x}_i - b) - E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\}] \,. \tag{A.26}$$

By Lemma 5.6, the second term of (A.26) converges to 0 a.s., uniformly on $\mathbf{a}$ and b as $n \to \infty$. Regarding the first term we have

$$\frac{1}{n}\sum_{i \in J_n(t_2)} E\{\rho(\mathbf{a}'\mathbf{x}_i - b)\} = \frac{1}{n}\sum_{i \in J_n(t_2)} E\{\rho[\mathbf{a}'\boldsymbol{\varepsilon}_i + (\mathbf{a}'\mathbf{X}_i - b)]\}$$

$$\geq \frac{1}{n}\sum_{i \in J_n(t_2)} E\{\rho(\mathbf{a}'\boldsymbol{\varepsilon}_i + t_0)\} = \frac{n(t_2)}{n} E\{\rho(\mathbf{a}'\boldsymbol{\varepsilon} + t_0)\}$$

$$\geq E\{\rho(\mathbf{a}'_0\boldsymbol{\varepsilon})\} + \varepsilon \tag{A.27}$$

for large n.

On the other hand, by the SLLN

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\hat{\mathbf{a}}'_n\mathbf{x}_i - \hat{b}_n) \leq \frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{a}'_0\mathbf{x}_i - b_0) = \frac{1}{n}\sum_{i=1}^{n} \rho(\mathbf{a}'_0\boldsymbol{\varepsilon}_i) \to E\{\rho(\mathbf{a}'_0\boldsymbol{\varepsilon})\} \quad \text{a.s.} \tag{A.28}$$

Now the lemma follows from (A.27) and (A.28).

Proof of Lemma 5.8:

Since ρ is bounded   is pointwise bounded. We will show now that   is equicontinuous.

Let $\varepsilon > 0$ be fixed. Let $\delta > 0$ be such that $|t - \tilde{t}| < \delta$ implies $|\rho(t) - \rho(\tilde{t})| < \varepsilon/2$. Let $\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \mathbf{C}$ be such that $\|\mathbf{a} - \tilde{\mathbf{a}}\| + |b - \tilde{b}| < \Delta$, for some $\Delta > 0$ to be chosen later.

By (ii.a), for any $K > 0$ we have

$$|h_n(\boldsymbol{\theta}) - h_n(\tilde{\boldsymbol{\theta}})| = \frac{1}{n}\left|\sum_{i=1}^{n} [E\{\rho(\mathbf{a}'\boldsymbol{\varepsilon}_i + \mathbf{a}'\mathbf{X}_i - b)\} - E\{\rho(\tilde{\mathbf{a}}'\boldsymbol{\varepsilon}_i + \tilde{\mathbf{a}}'\mathbf{X}_i - \tilde{b})\}]\right|$$

$$= \frac{1}{n}\left|\sum_{i=1}^{n} E\{\rho(Y + \mathbf{a}'\mathbf{X}_i - b) - \rho(Y + \tilde{\mathbf{a}}\mathbf{X}_i - b)\}\right|$$

$$\leq \frac{1}{n}\sum_{i \in J_n(K)} E\{|\rho(Y + \mathbf{a}'\mathbf{X}_i - b) - \rho(Y + \tilde{\mathbf{a}}'\mathbf{X}_i - \tilde{b})|\} +$$

$$+ \frac{1}{n}\sum_{i \notin J_n(K)} E\{|\rho(Y + \mathbf{a}'\mathbf{X}_i - b) - \rho(Y + \tilde{\mathbf{a}}\mathbf{X}_i - \tilde{b})|\} \qquad \text{(A.29)}$$

where $Y = \mathbf{a}'\boldsymbol{\varepsilon}$. By (ix), we can choose K such that $2(1 - n(K)/n)M < \varepsilon/2$ for all $n \geq N_\varepsilon$. We can now choose $\Delta > 0$ such that $2K\Delta < \varepsilon/2$. For this choice of K and Δ it follows that, for all $M \geq N_\varepsilon$

$$|h_n(\boldsymbol{\theta}) - h_n(\tilde{\boldsymbol{\theta}})| \leq \frac{\varepsilon}{2} = \varepsilon\ . \qquad \text{(A.30)}$$

This proves the lemma; since $h_n(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ for each fixed n, the cases $n \in N_\varepsilon$ are immaterial to the proof.

Proof of Lemma 5.9:

Let $\boldsymbol{\theta} = (\mathbf{a}, b) \neq \boldsymbol{\theta}_0$. By (x) there exists $N = N(\boldsymbol{\theta}) \in \mathbb{N}$, $\varepsilon = \varepsilon(\boldsymbol{\theta}) > 0$, and $\delta = \delta(\boldsymbol{\theta}) > 0$ such that for all $n \geqq N$

$$\frac{m(\varepsilon)}{n} = \frac{\{i : |\mathbf{a}'\mathbf{X}_i - b| \geq \varepsilon\}}{n} \geqq \delta \ . \tag{A.31}$$

By Lemma 5.1, there exists $\delta' > 0$ such that $E\{\rho(Y+\varepsilon)\} - E\{\rho(Y)\} = \delta' > 0$.

Now,

$$\begin{aligned} h_n(\boldsymbol{\theta}) &= \frac{1}{n} \sum_{i=1}^{n} E\{\rho[Y + (\mathbf{a}'\mathbf{X}_i - b)]\} = \\ &= \frac{1}{n} \left[\sum_{i \in I_n(\varepsilon)} E\{\rho[Y + (\mathbf{a}'\mathbf{X}_i - b)]\} + \sum_{i \in I_n(\varepsilon)} E\{\rho[Y + (\mathbf{a}'\mathbf{X}_i - b)] \right] \\ &\geqq \frac{1}{n} \left[\sum_{i \in I_n(\varepsilon)} (E\{\rho(Y)\} + \delta') + \sum_{i \in I_n(\varepsilon)} E\{\rho(Y)\} \right] \\ &= h_n(\boldsymbol{\theta}_0) + \frac{m(\varepsilon)}{n} \delta' \geqq h_n(\boldsymbol{\theta}_0) + \delta\delta' \ . \end{aligned} \tag{A.32}$$

Therefore, for $n \geqq N$ we have

$$h_n(\boldsymbol{\theta}) - h_n(\boldsymbol{\theta}_0) \geqq \delta\delta' > 0 \tag{A.33}$$

proving the lemma.

References

Anderson, T.W. (1984). "Estimating linear statistical relationships." *The Annals of Statistics*, **12**, No. 1, 1-45.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

Brown, M. (1982). "Robust line estimation with errors in both variables." *Journal of the American Statistical Association*, **77**, 377, Theor. and Methods Section.

Carroll, R. and Gallo, P. (1982). "Some aspects of robustness in the functional errors-in-variables regression model." *Comm. Statist. Theor. Meth.* **11** (22), 2573-2585.

Chung, K.L. (1974). *A Course in Probability Theory*. Academic Press, New York.

Fuller, W. (1980). "Properties of some estimators for the error-in-variables model." *The Annals of Statistics*, **8**, No. 2, 407-422.

Gleser, L. (1981). "Estimation in a multivariate error in variables regression model: large sample results." *The Annals of Statistics*, **9**, No. 1, 24-44.

Hampel, F.R. (1968). "Contributions to the theory of robust estimation." Ph.D. Thesis, University of California, Berkeley.

Hampel, F.R. (1971). "A general qualitative definition of robustness." *Ann. Math. Stat.*, **42**, 1887-1896.

Hampel, F.R. (1974). "The influence curve and its role in robust estimation." *J. Amer. Statist. Assoc.*, **62**, 1179-1186.

Huber, P.J. (1964). "Robust estimation of a location parameter." *Ann. Math. Statist.*, **35**, 73-101.

Huber, P.J. (1967). "The behavior of maximum likelihood estimates under non-standard conditions." *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1. University of California Press, Berkeley.

Huber, P.J. (1973). "Robust regression: asymptotics, conjectures and Monte Carlo." *Ann. Statist.*, **1**, 799-821.

Huber, P.J. (1981). *Robust Statistics,* John Wiley & Sons, New York.

Kelly, G. (1984). "The influence function in the errors in variables problem." *The Annals of Statistics,* **12**, No. 1, 87-100.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics,* Vol. 2, Fourth Edition. Griffin, London.

Madansky, A. (1959). "The fitting of straight lines when both variables are subject to error." *J. Amer. Statist. Assoc.*, **54**, 173-205.

Moran, P.A.P. (1971). "Estimating structural and functional relationships." *Journal of Multivariate Analysis,* **1**, 232-255.

Reiersol, O. (1950). "Identificability of a linear relation between variables which are subject to error." *Econometrica,* **18**, 375-389.

END
FILMED
3-86
DTIC